

STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 98891

TO: Maher Haddad
Location: 9d08 / 9e12
Wednesday, July 16, 2003
Art Unit: 1644
Phone: 306-3472
Serial Number: 09 / 647544

From: Jan Delaval
Location: Biotech-Chem Library
CM1-1E07
Phone: 308-4498

jan.delaval@uspto.gov

Search Notes

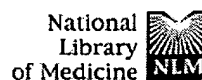
- ① Page # 9, Line 27, Need seq ID No.
② Page 15, check Amed By
18 4 4 65
3 30, Table I Need seq 4

New matter
seq ID No.
11
==

Examiner's notes

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 -- 703-308-4498
jan.delaval@uspto.gov

THIS PAGE BLANK (USPTO)



PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	OMIM	Books	
Search	PubMed	▼	for					Go	Clear
Limits		Preview/Index		History		Clipboard		Details	

Display	Abstract	▼	Show:	20	▼	Sort	▼	Send to	Text	▼
---------	----------	---	-------	----	---	------	---	---------	------	---

1: Cytogenet Cell Genet. 1999;87(3-4):238-44.

[Related Articles, Links](#)

Entrez
PubMed



The integrin alpha10 subunit: expression pattern, partial gene structure, and chromosomal localization.

Lehnert K, Ni J, Leung E, Gough S, Morris CM, Liu D, Wang SX, Langley R, Krissansen GW.

PubMed
Services

Department of Molecular Medicine, Faculty of Medicine and Health Science, The University of Auckland, Auckland, New Zealand.

Related
Resources

Herein we report the cloning of cDNAs and incompletely processed hnRNAs from endothelia and heart that encode the alpha10 subunit forming part of the novel collagen type II-binding integrin alpha10beta1 of chondrocytes. Analysis of hnRNA clones and reported expressed sequence tags revealed the positions of 17 putative intron-exon splice junctions shared with those of the p150,95 (ITGAX) gene. Human alpha10 transcripts of 5.4 and 1.8 kb were not restricted to chondrocytes but, instead, were widely expressed in a panel of 24 tissue types, where the highest expression was found in muscle and heart. The human alpha10 subunit gene (ITGA10) was localized to band q21 of chromosome 1. Copyright 2000 S. Karger AG, Basel

PMID: 10702680 [PubMed - indexed for MEDLINE]

Display	Abstract	▼	Show:	20	▼	Sort	▼	Send to	Text	▼
---------	----------	---	-------	----	---	------	---	---------	------	---

[Write to the Help Desk](#)

[NCBI](#) | [NLM](#) | [NIH](#)

[Department of Health & Human Services](#)

[Freedom of Information Act](#) | [Disclaimer](#)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 30.365 Seconds

(without alignments)
1594.036 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106
Sequence: 1 MELPVTHTLFLPLVFLTGLC.....GFPAHKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6106	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	2429	39.8	1189	1 ITAH_HUMAN	O94ux5 homo sapien
3	1894.5	31.0	1180	1 ITAI_RAT	P18614 rattus norv
4	1887	30.9	1151	1 ITA1_HUMAN	P56199 homo sapien
5	1756	28.8	1181	1 ITA2_HUMAN	P17301 homo sapien
6	1754	28.7	1170	1 ITA2_BOVIN	P53710 bos taurus
7	1744	28.6	1178	1 ITA2_MOUSE	O62469 mus musculu
8	1131.5	18.5	1162	1 ITAD_HUMAN	Q13349 homo sapien
9	1105	18.1	1170	1 ITAL_HUMAN	P24063 mus musculu
10	1097	18.0	1163	1 ITAM_MOUSE	P20701 homo sapien
11	1079	17.7	1152	1 ITAN_HUMAN	P11215 homo sapien
12	1051	17.2	1163	1 ITAX_HUMAN	P20702 homo sapien
13	1022.5	16.7	1153	1 ITAM_MOUSE	P05655 mus musculu
14	953	15.6	1167	1 ITAE_MOUSE	O60677 mus musculu
15	932.5	15.3	1179	1 ITAE_HUMAN	P38570 homo sapien
16	738	12.1	1135	1 ITA9_HUMAN	Q13797 homo sapien
17	690	11.3	1039	1 ITA4_MOUSE	O00651 mus musculu
18	689.5	11.3	1032	1 ITA4_XENLA	O91687 xenopus lae
19	684	11.2	1038	1 ITA4_HUMAN	P13612 homo sapien
20	674.5	11.0	285	1 ITA1_CHICK	O90615 gallus galli
21	629.5	10.3	1181	1 ITA7_HUMAN	Q13683 homo sapien
22	602.5	9.9	1130	1 ITA6_HUMAN	P23209 homo sapien
23	600.5	9.8	1066	1 ITA3_HUMAN	P26006 homo sapien
24	596.5	9.8	1072	1 ITA6_CHICK	P26007 gallus galli
25	596.5	9.8	1179	1 ITA7_MOUSE	O61738 mus musculu
26	586.5	9.6	1050	1 ITA5_XENLA	O06174 xenopus lae
27	586	9.6	1044	1 ITA6_MOUSE	O61739 mus musculu
28	584.5	9.6	1041	1 ITA6_MOUSE	P43406 mus musculu
29	581.5	9.5	1053	1 ITA3_MOUSE	O62470 mus musculu
30	576.5	9.4	1044	1 ITA8_CHICK	P26009 gallus galli
31	574	9.4	1048	1 ITA9_HUMAN	P06156 homo sapien
32	573.5	9.4	1106	1 ITA7_RAT	O63358 rattus norv
33	572.5	9.4	1066	1 ITA3_CRISP	P17852 cricetidae

34	566	9.3	1053	1 ITA5_MOUSE	P11688 mus musculu
35	563.5	9.2	1034	1 ITA9_CHICK	P26008 gallus galli
36	562	9.2	1049	1 ITA5_HUMAN	P08648 homo sapien
37	541	8.9	1146	1 ITA1_DROME	O24247 drosophila
38	540.5	8.9	1025	1 ITA8_HUMAN	P53708 homo sapien
39	534	8.7	1039	1 ITAB_HUMAN	P08514 homo sapien
40	504	8.3	1033	1 ITAB_MOUSE	O94um0 mus musculu
41	485	7.9	1226	1 ITA2_MOUSE	P34446 caenorhabdi
42	475	7.8	1396	1 ITA2_DROME	P12080 drosophila
43	447.5	7.3	1000	1 ITA5_DROME	O9w1m8 drosophila
44	445.5	7.3	1139	1 ITA1_CAEEL	O03600 caenorhabdi
45	422.5	6.9	1115	1 ITA3_DROME	O44386 drosophila

ALIGNMENTS

```

RESULT 1
ID ITAG_HUMAN STANDARD; PRT: 1167 AA.
AC O75578; Q9UH28;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-10 precursor.
GN ITGA10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Articular chondrocytes;
RX MEDLINE=98352078; PubMed=9685391;
RA Camper L., Hellman U., Lundgren-Akerlund E.;
RT "Isolation, cloning, and sequence analysis of the integrin subunit
RT alpha10, a beta1-associated collagen binding integrin expressed on
RT chondrocytes."
RL J. Biol. Chem. 273:20383-20389 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart, and Endothelial cells;
RX MEDLINE=20169197; PubMed=10702680;
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
RA Wang S.-X., Langley R., Krissansen G.W.;
RT "The integrin alpha10 subunit: expression pattern, partial gene
RT structure, and chromosomal localization."
RL Cytogenet. Cell Genet. 87:238-244 (1999).
CC - FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC ASSOCIATES WITH BETA-1.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.
CC - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC - SIMILARITY: CONTRAINS 1 VWFA DOMAIN.
CC - SIMILARITY: CONTRAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF074015; AAC31952.1; -
DR EMBL: AF112345; AAP21944.1; -
DR EMBL: AF172723; AAF61638.1; -
DR HSSP: P17301; IAOX.
DR Genew: HGNC:6135; ITGA10.

```

QY	30	GIIVLGHYLRRODRDSSFLAREIRTAASDPDRPFENVTDALDIDYDALGDSIFGLEGS	360
Db	301	GIIVLGHYLRRODRDSSFLAREIRTAASDPDRPFENVTDALDIDYDALGDSIFGLEGS	360
QY	361	HAENESSFGLIEMSOIGFSTHRLKDGLFGVAVYDWGGSVLMTEGHRLEPPMALDEF	420
Db	361	HAENESSFGLIEMSOIGFSTHRLKDGLFGVAVYDWGGSVLMTEGHRLEPPMALDEF	420
QY	421	PPALQNHAAVLYGSVSSMLRGRLFLSAGPRRRHGXVYAFOLKKDGAVRVAOSIQE	480
Db	421	PPALQNHAAVLYGSVSSMLRGRLFLSAGPRRRHGXVYAFOLKKDGAVRVAOSIQE	480
QY	481	QIASYFSGELCPDIDBDGTDVLLVAAPMLGCONKETGVVYVYVQOSSLTLOGTLQ	540
Db	481	QIASYFSGELCPDIDBDGTDVLLVAAPMLGCONKETGVVYVYVQOSSLTLOGTLQ	540
QY	541	PEPPDARFGFMAGLPDLNDGFADVAGPRLDGQAGALYVHGQSIVRRHPRORIA	600
Db	541	PEPPDARFGFMAGLPDLNDGFADVAGPRLDGQAGALYVHGQSIVRRHPRORIA	600
QY	601	AAMEPALSYBERSVDGSLDLDGDIYDVAGAGAAIILSSRPVHLYTSLEVTPOAIS	660
Db	601	AAMEPALSYBERSVDGSLDLDGDIYDVAGAGAAIILSSRPVHLYTSLEVTPOAIS	660
QY	661	VVGRDRRRGCGAVCLTAALCFQVTSRTPGHMHQFMRFTASLEWTAGARAPGSGQ	720
Db	661	VVGRDRRRGCGAVCLTAALCFQVTSRTPGHMHQFMRFTASLEWTAGARAPGSGQ	720
QY	721	RUSPRRLASVGNVTCBQLHPVLDTSYLRLPVALYTFALDNTTKGPGVLNCSPTIQ	780
Db	721	RUSPRRLASVGNVTCBQLHPVLDTSYLRLPVALYTFALDNTTKGPGVLNCSPTIQ	780
QY	781	KLVPFKDCGDPNECVTDLVYQVMDIDGSKAPFVVRGGRRKVLVSTLENKENAYNT	840
Db	781	KLVPFKDCGDPNECVTDLVYQVMDIDGSKAPFVVRGGRRKVLVSTLENKENAYNT	840
QY	841	SUSIIFSRMLHASTLPQRESPIKVECAPSAHARLCSVGHPPVQTAKVTLLEEFSC	900
Db	841	SUSIIFSRMLHASTLPQRESPIKVECAPSAHARLCSVGHPPVQTAKVTLLEEFSC	900
QY	901	SSLSLVSOFCKLTASSDSLEBNGTLOENTAOISAVIOYEPHLFSSBSTLHRYEVHPYGL	960
Db	901	SSLSLVSOFCKLTASSDSLEBNGTLOENTAOISAVIOYEPHLFSSBSTLHRYEVHPYGL	960
QY	961	PVGPGEFFTLTRYONLCGYVSGIILSALLPAVHAGNFFLSIOYITNNAACIQONT	1020
Db	961	PVGPGEFFTLTRYONLCGYVSGIILSALLPAVHAGNFFLSIOYITNNAACIQONT	1020
QY	1021	EPGPFPVHBEELQHTNRUNGNSNTQOAVRCHLGOLAKTEVSGLRLVHNEFFRABAKR	1080
Db	1021	EPGPFPVHBEELQHTNRUNGNSNTQOAVRCHLGOLAKTEVSGLRLVHNEFFRABAKR	1080
QY	1081	SILTUVSTFELGIBESBVQLTEASWSSESLLEVQTRPILISLMLIGSVLGGLLALL	1140
Db	1081	SILTUVSTFELGIBESBVQLTEASWSSESLLEVQTRPILISLMLIGSVLGGLLALL	1140
QY	1141	VFCIMKLGFFAHKKIPEEKKEKLEQ	1167
Db	1141	VFCIMKLGFFAHKKIPEEKKEKLEQ	1167
RESULT 2			
ITAH_HUMAN			
ID	_ITAH_HUMAN	STANDARD;	PRT; 1189 AA.
AC	Q9UKX5; O9UKO1;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Integrin alpha-11 precursor.		
GN	ITGAL1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal heart, and Osteoblast;
 RX MEDLINE=99417678; PubMed=10486209;
 RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
 RA Wang S.-X., Morris C.M., Kirsansen G.W.;
 RT "Cloning, sequence analysis, and chromosomal localization of the novel
 RT human integrin alpha1 subunit (ITGA11).";
 RL Genomics 60:179-187(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal muscle, and Uterus;
 RX MEDLINE=99395147; PubMed=10464311;
 RA Velling T., Kuesche-Gullberg M., Sejersten T., Gullberg D.;
 RT "cDNA Cloning and Chromosomal Localization of Human Alpha(11)
 RT Integrin. A collagen-binding, 1 domain-containing, beta(1)-associated
 RT integrin alpha-chain present in muscle tissues.";
 RL J. Biol. Chem. 274:25735-25742(1999).
 RN [3]
 RP SEQUENCE OF 954-1188 FROM N.A.
 RC TISSUE=Fibroblast;
 RA Andreu N., Escarceller M., Sumoy L.;
 RL Submitted (Jun-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
 CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
 CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
 CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
 CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
 CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
 CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
 CC FETAL MUSCLE CELLS (IN VITRO).
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 WFPA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF109681; AAF01258.1; -;
 DR EMBL, AF137378; AAD51919.2; -;
 DR EMBL, AL359064; CAB94392.1; -;
 DR HSSP, P17301; IAOX.
 DR Genew; HGNC:6136; ITGA11.
 DR MIM: 604789; -;
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; WVF_A.
 DR Pfam; PF00092; wva; 1.
 DR Pfam; PF01639; FG-GAP; 5.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; WVFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; WVA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 DR PROSITE; PS50234; WVFA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Repeat; Calcium; Magnesium; Polymorphism.
 FT SIGNAL 1 22
 FT CHAIN 23 1189 INTEGRIN ALPHA-11.
 FT DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1143 1165 POTENTIAL.

FT DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 38 94 FG-GAP 1.
 FT REPEAT 102 163 FG-GAP 2.
 FT DOMAIN 167 345 WVF_A.
 FT REPEAT 359 420 FG-GAP 3.
 FT REPEAT 422 475 FG-GAP 4.
 FT REPEAT 477 537 FG-GAP 5.
 FT REPEAT 539 598 FG-GAP 6.
 FT REPEAT 601 653 FG-GAP 7.
 FT DOMAIN 1154 1162 POLY-LEU.
 FT DOMAIN 1174 1177 POLY-ARG.
 FT CA_BIND 488 496 POTENTIAL.
 FT CA_BIND 551 559 POTENTIAL.
 FT CA_BIND 613 621 POTENTIAL.
 FT DISULFID 76 83 BY SIMILARITY.
 FT DISULFID 121 139 POTENTIAL.
 FT DISULFID 129 159 POTENTIAL.
 FT DISULFID 659 668 BY SIMILARITY.
 FT DISULFID 674 729 BY SIMILARITY.
 FT DISULFID 781 787 BY SIMILARITY.
 FT DISULFID 881 893 BY SIMILARITY.
 FT CARBOHYD 82 82 N-LINKED (GLCNAC...)
 FT CARBOHYD 95 95 N-LINKED (GLCNAC...)
 FT CARBOHYD 291 291 N-LINKED (GLCNAC...)
 FT CARBOHYD 311 331 N-LINKED (GLCNAC...)
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...)
 FT CARBOHYD 449 449 N-LINKED (GLCNAC...)
 FT CARBOHYD 462 462 N-LINKED (GLCNAC...)
 FT CARBOHYD 528 528 N-LINKED (GLCNAC...)
 FT CARBOHYD 642 642 N-LINKED (GLCNAC...)
 FT CARBOHYD 694 694 N-LINKED (GLCNAC...)
 FT CARBOHYD 857 857 N-LINKED (GLCNAC...)
 FT CARBOHYD 894 894 N-LINKED (GLCNAC...)
 FT CARBOHYD 973 973 N-LINKED (GLCNAC...)
 FT CARBOHYD 1032 1032 N-LINKED (GLCNAC...)
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC...)
 FT VARIANT 433 433 V -> M.
 FT VARIANT 524 524 /FTID=VAR_009889.
 FT VARIANT 524 524 R -> L.
 FT VARIANT 972 972 /FTID=VAR_009890.
 FT VARIANT 1003 1003 L -> P.
 FT VARIANT 1030 1030 /FTID=VAR_009891.
 FT VARIANT 1030 1030 I -> M.
 FT VARIANT 1094 1094 /FTID=VAR_009892.
 FT VARIANT 1094 1094 MISSING.
 FT VARIANT 1094 1094 /FTID=VAR_009893.
 FT VARIANT 1094 1094 L -> V.
 FT VARIANT 1094 1094 /FTID=VAR_009894.
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08A4A4CD52 CRC64;
 Query Match 39.8%; Score 2429; DB 1; Length 1189;
 Best Local Similarity 43.1%; Pred. No. 5.7e-167;
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;
 QY 1 MELPFVTHFLPLVFLTGLCSFNLDEHHPRLFGPPPAERGVSYLQHVGGGQRMVLVGA 60
 DB 1 MDLPRLGLVAVNALSLMPGFDTFPMNDTRKPRVIESRSRAFFGYTQGDIGSKMLVVGA 60
 QY 61 FWDGSGDRRDGYTRCPYGAHANAPCAKGLGDYOLGNSSHPAYNMHGLMSLFTDGGCG 120
 DB 61 PLETWGYOKTGDVYKCPV---IHGCTKLNIGRVTLSNVSERKMKMRIGLSLATNPKN 117
 QY 121 FMACAPLMSRAGSSVFSGGICARVDASFQOGSLAPTAORCPFTMDVTVYLDGNSITYP 180
 DB 118 FLACSPLMSSHEGSSYTTGMCISVNSFRSKVYAPLQRCQTYMDIVYLDGNSITYP 177
 QY 181 WSEVOTFLRLRVGLKFLIDPEQIOVGLVOYGSVPHWMSLADFPRTKEEVVRAKMLSRREG 240
 DB 178 WVEVGHFLINILTKRYIGPQIOYGVGDVHVEFLINDYRSKVDVVEASHIEGRG 237
 QY 241 RETKRAQIMVACTEGESQSGRPEARLLVVTVDGSHGEBELPAALKACEAGRVTRY 300
 DB 238 TETRTAFIEIPARSEAFQK--GGRKAKKVMIVITDGHSDSPDLEKVIQOOSERDVTTRY 295

QY 301 GIAVLGHYLRORPSSFLREIRTIASDPDERFFNVNTEAALTDIADLRIFGLGEGS 360
 DB 296 AVAVLGYNRRGINPFTFLNEIKYIASDPDCKFFNVNTEAALTDIADLRIFGLGEGS 355
 QY 361 HAENESSFGLGEGSIOGFSTRHLKDGILFGVNAVYDMGVSYLEGHRLEPPMALEDEF 420
 DB 356 N-KNETSFGLGEGSIOGFSTRHLKDGILFGVNAVYDMGVSYLEGHRLEPPMALEDEF 414
 QY 421 PPAALNHAAYLVGSVSMILRGSRRLFLSGAPRRHRGKVIATOLKDDGVAVRVAOSLOOE 480
 DB 415 PEEELKNGHAGLVGVTSVSSVSSROGRVYVAGAPRNHGVKYLETMMNNSSLTTHQMLKRG 474
 QY 481 QIGSYGSELCPIEDTRDGTDDVLLVAAPFLCPQNKETGRVYVYVVGQOSLLTLLOQTQ 540
 DB 475 QIGSYGSEITSDIDGDGTVDLLVGAPEYFN-EGREGRKVVYVEL-RQNRVYVNGTLK 532
 QY 541 PEPP-ODARFGFAMGALPDUNOGFADVAVGAPLEDGHOALYLYHGTOSGVAPHPAOKR 599
 DB 533 DSHSYGNARFGSSIASVRDLNODSYNDVVVGALEENHAGAIYIFGFGRSILKTEKOR 592
 QY 600 AASMHALSYFGRSVDRDLDDDLVDVAVAGOGAAILSSRPVHLTPSLVTPPOAI 659
 DB 593 TASELATGLQYFCSJHGQDLNEDGLIDAVGALGNVAVLMSRPVAVQINASLHPEPSKI 652
 QY 660 SVVQDRCRRGQCAVCLTAALCFQVTSRTPGRMDHOFYMERFTASLDEMTAGARAAPDGG 719
 DB 653 NIFHRCKSGRATCLTAALFCTPFLFAPHPOTTVGIRYVNTMDERRTPPAHLDEG 712
 QY 720 QRISSPRLRLSVGNVTCQOLHFLVLDTSYLRYVALTVTFALDNTTKPGVNLGSPST 779
 DB 713 DRETNPAVLSSQCELCERINFVLDYADVKPFTSVESLEDDPH-GPMLDDGMPTL 771
 QY 780 QKVPSPKXCGPNECVTDLVLOVNDI-----RGSXK-----APFV 817
 DB 772 RVSVPTWNGCNEDEHCPPDLVLDARSDDLPTAMEYCOGLKRPQDCSANTLSFTTVFII 831
 QY 818 RGRGRKVLVSTLTENKENVNTSLIIFSRNLHLASLTPORESPIKVECAAPS-AHAR 875
 DB 832 ESTROKVAVEATLENNGENAVSTVLNISGSANQFASLQKEDSDSIECVNERLQOQ 891
 QY 876 LGSVGHVPTQAKVTFLEFEFSCSSLSQVFGKLTASSDSIERNGTLOENTAOQSANI 935
 DB 892 VCVNVSYPFRACKKVFRLDFEFSKSIFLHLEIEILAGSDSNERSTKEDNVAAPRFLH 951
 QY 936 QVEPHLLFSESTLHRYEYVPGTLP--VGPGEFPTTLRVONLGCYVVSGLISALLPA 993
 DB 952 KTEADVLFTRSSLSHYEVKLSLERYDGIQPFSCIFRIQNLGLFPIHGLIMKITTPI 1011
 QY 994 VAHGVNVLSTLSQVITN--NASC-IVONLTPPPGPVHPEELOHTNRLNGSNTQCOVR 1049
 DB 1012 ATMSGRLKLRLFLDEVANTSCNIMGNSTEVPRPVE-EDLRARPOLNHSNDVVSIN 1070
 QY 1050 CHLGOLAKGTEVSGLLRLVNEHFFRPAKFKSLTVVSTELGTEBSVYLQLTASWSS 1109
 DB 1071 CNI-RLVPONEINFLHGLNMLRSLKALKKXKIMVNAALQCFHSPTFREEDPSROI 1129
 QY 1110 LLEVOTRPLILSLMLIGSVLGLLLALVLCMLKGFPAKFKIPEEKREKLE 1166
 DB 1130 VBEISKOEDQVPIWITVSTLGLLLALLVLAALKGFPSAR---RRREPGD 1182

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OC NCBI_Taxid=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90338125; PubMed=2380249;
 RA Ignatius M.J., Houde T.H., Large T.H., Barton A.,
 RA Esch F., Carbonetto S., Reichardt L.F.;
 RT "Molecular cloning of the rat integrin alpha 1-subunit: a receptor
 RT for laminin and collagen";
 RL J. Cell Biol. 111:709-720 (1990).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
 RX MEDLINE=99313197; PubMed=10386626;
 RA Nolte M., Pepinsky R.B., Vanyaminov S.Y., Kotellianeky V.,
 RA Gotwals P.J., Kapusnas M.;
 RT "Crystal structure of the alphabeta1 integrin I-domain: insights into
 RT integrin I-domain function";
 RL FEBS Lett. 452:379-385 (1999).
 CC -I- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
 CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
 CC E-R IN COLLAGEN.
 CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
 CC ASSOCIATES WITH BETA-1.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52140; CAA36384.1; -;
 DR PIR: A35854; A35854.
 DR HSP: P17301; IAOX.
 DR InterPro: IPR000413; Integrin_alpha.
 DR InterPro: IPR002035; VWF_A.
 DR Pfam: PF00092; vwa; 1.
 DR Pfam: PF00357; Integrin_A; 1.
 DR Pfam: PF01839; FG-GAP_5.
 DR PRINTS: PR00453; VWFADOMAIN.
 DR SMART: SM00191; Int_alpha; 5.
 DR SMART: SM00327; vwa; 1.
 DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE: PS50234; VWFA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Repeat; Calcium; Magnesium.
 FT SIGNAL 1 28
 FT CHAIN 29 1180 INTEGRIN ALPHA-1.
 FT DOMAIN 29 1142 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1143 1165 POTENTIAL.
 FT DOMAIN 1166 1180 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 44 103 FG-GAP 1.
 FT REPEAT ? ? FG-GAP 2.
 FT DOMAIN 175 388 VWFA.
 FT REPEAT 377 432 FG-GAP 3.
 FT REPEAT 433 484 FG-GAP 4.
 FT REPEAT 485 565 FG-GAP 5.
 FT REPEAT 567 626 FG-GAP 6.
 FT REPEAT 629 681 FG-GAP 7.
 FT CA_BIND 497 505 POTENTIAL.
 FT CA_BIND 579 587 POTENTIAL.
 FT CA_BIND 641 649 POTENTIAL.
 FT SITE 1168 1172 GPRK MOTIF.
 FT DISULFID 82 92 BY SIMILARITY.
 FT DISULFID 687 696 BY SIMILARITY.

```

FT DISULFID 702 755 BY SIMILARITY.
FT DISULFID 807 813 BY SIMILARITY.
FT DISULFID 877 885 BY SIMILARITY.
FT DISULFID 1029 1062 BY SIMILARITY.
FT DISULFID 1066 1073 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1180 AA; 130808 MW; 855DA2BE02362EE4 CRC64;

```

```

Query Match 31.0%; Score 1894.5; DB 1; Length 1180;
Best Local Similarity 36.5%; Pred. No. 2, 1e-128;
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;

```

```

QY 13 LVFLTGLCSPPNLDHHPRLFPQPPAEFGYVLQHVGGGRMVLGVAEWDSPSGDRGD 72
DB 19 LTVIILGFCVFNVDVNSMFSFGPVEDMFGYTVQCENEGKMWLIGSLVPGQPKARATGD 78
QY 73 VYRCVGGAHNAPCAKKGHIGDYOLGNSHPAV-----NHLGMSLLETTDGGGPMACPL 127
DB 79 VKCEVGRERAMPVCKLDP---VNTSLPNVTEIKENMTFG-STLVNPNNGFLACGL 133
QY 128 WSRACGSSVSSGICARVDASFOPQSLAPTACPTVMVIVLVDGNSIYPMSEVQTF 187
DB 134 VYRCGHLHTTGICSDVSPFTQVNSFAP-VQECSTQIDYIVLVDGNSIYPMSEVIAF 192
QY 188 LRLVGLKFLIDPQIQVGLVQGESFVHWSLGFPTKEEVVRAAKNLSRRGRETAKQ 247
DB 193 LINDLKRMIDIGKQOVGIVQGENVTHEFNINKYSSTEBEVLVAANKIGRQGLQTMFAL 252
QY 248 AIMVACTEGFSOSHGRPEAARLLVVTTCGESHDEELPAALKAEEAGVNTYGVAVLGH 307
DB 253 GIDTARKEFTARGARRGVKKVMVLTVDGESHDNVRLKQVQIDCEDENIQFSSIALIGH 312
QY 308 YLRRODSSPLREIRTIASDDERFFVNTDEALTIIDVALDGRIFGLSESHANESS 367
DB 313 YNRGNLSTKEFVEELIKSISTEPTKEHFVNSDELALVTIVKALGRIFPLEXTADQSAAS 372
QY 368 FGLEMSQIGFSTHRLKDGILFPMVAGAYDWGSSVLMLEGGHRLFPPMALEDEFPALONH 427
DB 373 FEMEMSQIGFSAHYQSDWMLGAVGAYDNGITVMQKANOVI PHNTTQTE--PAKME 430
QY 428 --AALVGSVSSMLRGGRRLFLSGAPRRHKGKVIAPOLKDGAVRVAQSIQSGQIGSY 485
DB 431 PLASYIGYVNSATIPGD-VLYIAGOPRNHNGOVVITM--BDGNINILQTLGGQIGSY 488
QY 486 FGSSELCPDITDGTDLVLAAPMFLGPNKQETGRVYVLYVGOOSLITLQGTLP---- 541
DB 489 FGSVLTIDIDKSYDILLVGAAPMYMGTKEEGQGVYVYAN-NQRFYQMSLEPIRT 547
QY 542 -----EPQDARFGFMALPDINODGFADVAVGALEDDHOGALYLYHG 586

```

```

DB 548 CSSLKDNSCTKENKNEBCGARFGTALAAYKDLNVDFENVYIGALEDDHAGAVIYHG 607
QY 587 TQSGVRPAPORIAAASMPHALSYGRSVDRDLDDGDLVDVAAGAGAAIILSRPIV 646
DB 608 SKGTRERAVQRIPEGGCKTLKFFGQIHGBMDLNGCLTDVYIGGAGALFVARVA 667
QY 647 HLTPELVTPQAIISVVRDCCRROGAECVLTALCFQYTSRTPGRMDHOFYRPTASDE 706
DB 668 VVKVTMNEPKKVNIOKKNCGVEGETYCINATMCFHKLKSKEDSIYEAADQYRVTLDS 727
QY 707 WTAGRAAFDQSGORLSPRLPLSVGNVTCQQLHNVLDTSYLRPVALYTFALDNTTK 766
DB 728 LRQISRSEFSGTOEKRIOR--NITYRESECRHSGFYMDKHDFODSVAVTIDF--NLTJ 782
QY 767 P--GPVNEGSPTSIOXLVPPSKDQCPNECVTDVLQVNMNDIGSRAPFVGRGRKV 824
DB 783 PENGEVLDALPNSVHEIIPRAKDCGKERCIISDITLVST---TEKSLIVSQHDF 838
QY 825 LVSTLENRKENAVYTSLSIIFSRNLHSLTP--QRESPIVECAAPSAHARLCSGHPV 883
DB 839 NVSLTVKKKGSAYVTRVVGHSPLVLSGIEIQKDS-----CESNQTICRGVYF 891
QY 884 FQTKAKTFLLEPFESGSLSYVFGKLTASDSIERNGTIOENTAOISAYIQYEPHLLF 943
DB 892 LRAGETVTFKIIIFONTSHLENAIINHLSATSDSEEPLESINDNEVNISIPKYEVGQF 951
QY 944 SSESTLHRYEHPYGTLP-----VGPPEKRTLRVONLCGVVSGLIISALP-AVAH 996
DB 952 YSSASEHHISVAANETIPEFINSTEDIGENINVFYTIKRGHPPELQLSISFNLTRAD 1011
QY 997 GGNVFLSLSOVITNNASCIYONLTPEPG-----PPVPEELQHTNRINGSNTOCOVVC 1050
DB 1012 GYPVLVYIGWSSSDVNRPRSLDEPFGINSKKTKISKEVLKRGTIQDSSSTGVATI 1071
QY 1051 HLGQLAKC-TEVSGCLRLVNEPFRRAKFSLYVSTFELCTEGSVLQLTASRWSES 1109
DB 1072 TCSLPSDLSQVNSL-LMKPPTIRAHFESLNLTLTGEKSEKNS-LTSSSRKREL 1128
QY 1110 LLEVQVTRPLILSILMILIGSVGLLALLLPCMLKGFPAHKIPEEKREE 1163
DB 1129 AIQISKDGLPGRVPLWTLISAFGLLMLLILALMKIGFF--KRLPKKMEK 1180

RESULT 4
ID ITA1_HUMAN STANDARD; PRT; 1151 AA.
AC P56199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN ITGA1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=93155124; PubMed=8428973;
RA Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT "Expression of native and truncated forms of the human integrin alpha
RT 1 subunit.";
RL J. Biol. Chem. 268:2989-2996 (1993).
CC -!- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC E-R IN COLLAGEN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 VWFA DOMAIN.

```

CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD49a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prot/cd/cd49a.htm".
DR HSSP: P17301; IAOX.
DR Genew; HGNC:6134; ITGA1.
DR MIM; 192968; .
DR InterPro: IPR000413; Integrin_alpha.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00092; vwa; 1.
DR Pfam: PF00357; Integrin_A; 1.
DR Pfam: PF01839; FG-GAP; 5.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS0234; WFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Repeat; Calcium; Magnesium.
FT DOMAIN 1 113 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 114 116 POTENTIAL.
FT DOMAIN 117 115 CYTOPLASMIC (POTENTIAL).
FT REPEAT 16 75 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT ? ? WFA.
FT DOMAIN 147 360 FG-GAP 3.
FT REPEAT 349 404 FG-GAP 4.
FT REPEAT 405 457 FG-GAP 5.
FT REPEAT 459 520 FG-GAP 6.
FT REPEAT 540 599 FG-GAP 7.
FT REPEAT 602 654 FG-GAP 7.
FT CA_BIND 470 478 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 614 622 POTENTIAL.
FT SITE 1139 1142 GEFKR MOTIF.
FT DISULFID 54 64 BY SIMILARITY.
FT DISULFID 660 669 BY SIMILARITY.
FT DISULFID 675 728 BY SIMILARITY.
FT DISULFID 780 786 BY SIMILARITY.
FT DISULFID 850 858 BY SIMILARITY.
FT DISULFID 1002 1034 BY SIMILARITY.
FT DISULFID 1037 1044 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 77 77 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 84 84 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 189 189 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 374 374 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 671 671 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 720 720 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 752 752 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 812 812 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 855 855 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 880 880 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 887 887 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 911 911 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 946 946 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 980 980 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 1074 1074 N-LINKED (GLCNAC; .) (POTENTIAL).
FT CARBOHYD 1085 1085 N-LINKED (GLCNAC; .) (POTENTIAL).
SQ SEQUENCE 1151 AA; 127837 MW; 683F3C1AABF52808 CRC64;

Query Match 30.9%; Score 1887; DB 1; Length 1151;
Best Local Similarity 36.3%; Pred. No. 7e-126;
Matches 459; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

23. FNIDHPRLLFGPRPEAFGYSVLQHVGGQGRWMLVGAWPDGSGDRGRDYYRCVPVGAH 82

Db 1 FNVVDKXNMTSGPVEDMFGTVQYQYENEBEGKWVLIGSPVVGQPKNRGQGVYKCPVGGGE 60
QY 83 NAFPCAKGHIGDYQLGNSSHPAV----NMHLGMSLLETDGCGFMACAPLMSRACGSSVF 137
Db 61 SLPCVYKLLP-----VNTSIPNTEVKENMTTG-STLVTPNPGGFLACCPLYAYRGCHLHY 115
QY 138 SSGICARVDASFPQCGSLAPTAQRCPYMDVYIVLDGNSNIYPAGEVOTFRLRYGLKFI 197
Db 116 TTGICSDVSPFTQVNVNLIAP-VQECSTQDVIYIVLDGNSNIYPMDSVTAFLNDLKRMDI 174
QY 198 DPEQIQVQLVOYGESPVHEWMSLGDPRFKEEYVRAAKNLSRREGRETQAQIMVACTEGF 257
Db 175 GPKQIQVQVQYQYGENVTHEFNLNKYSSTEEVLYAAKKIVQGGQRTMTALGDTARKEAF 234
QY 258 SOSHGCRPEARLLVYVTDGESHDEELPALKACEARVTRYGAVLGHYLRDRDSS 317
Db 235 TEARGARGGVKKWVIYVTDGESHDRHLKVIYQDEDEENIQRFSLIAGSYNRGLSTEK 294
QY 318 FLREIRTIASDPDERFFFNVTDEALTDIYDALGRIFGLSGSHANESSFGLEMSQIGF 377
Db 295 FVEEIKSIASEPTEKHFPNVSDDELALVTITLGRITALEPTADQSAASEMEMSQIGF 354
QY 378 STHRLKDGILFGWVAGAYDWGSLWLBGGHRLFPFRMALEDFPPALONHAAYIGYSVS 437
Db 355 SAHYSQDWVMVGAVAYDMNGTVWQKASQIIPRNTFNVESTKKNRPLASLYIGTVNS 414
QY 438 MLIRGRLLFLFSGARFRFRHGRKVIAPOLKKGAVRVAASLOGEQISFGSELCPDIDR 497
Db 415 ATASSGDVLYIAGQRYVHTQVLIYRM-EDGNKIILDTLGEQIGSYFGSILVTTDIDK 473
QY 498 DGTITVLVAAPMPFPGPONKETGRVYVYLVGQSLTLTQGTLPQ----- 541
Db 474 DSNTIILVGAAPMWTGTEKEQKYYVAL-NQRFYQMSLEPRKOTCCSSRQNSCTT 532
QY 542 --BPPODARFGFAGALPDINODGFADVAVGAPLEDGHOGLYLYHGTQSGVRPAPQR 598
Db 533 EKNRPPCARFGTALIAVKDLNLDGFNDIVIGAPLEDHGAVVYHYSGKTIKREYQR 592
QY 599 IAAAPMALSTFSGSVSGRLDLOGDDLVDAVAGQGAAILLSRPYHLRPSLEVTQA 658
Db 593 IPSGGDKTLKFFGGSIGEMDLNDGLTDVITGGLGGAALFWSDDAVVAVVYKMFENK 652
QY 659 ISVORDCRRRGOEAVCTLAALCFQVTSRTGPRMDHQPYMFSTASLDEMTAGARAFDGS 718
Db 653 VNIQKGNHMEKREVCINATVCEPVKLSKEDTITFADLQRYVLDSLRQISRFSGT 712
QY 719 GQRSLPRRLRSLVGNVTCQQLHFAVLDTSYLRLPVALTVPALDNTYKP--GVLNCGSP 776
Db 713 QERKVR--NITVRKSECTKHSFYMLDGHDFQDSVRITLDF--NLTPENGVPVLDLSLP 767
QY 777 TSIOQLVPSKDCGDNCEVVDVLYQVMMDIRGSKAFVVRGGRKYLVSITLENREN 836
Db 768 NSVHEIYFPADCGKKEKICISDLHV---ATTEKDLIYRSQNDKKNVSLTVNNTDS 823
QY 837 AYNTSLIIFSRNLHLASTPQRESPIKYECAASAHARLCSVGHVPQTGAKYFLEEF 896
Db 824 AYNTTIHYHSPNLVFSGI-----EAIQKQSC--ESNNHITCKYGVPIFARGMVFYKILF 877
QY 897 EFGSSLLSQVFGKLTASDSLERNGLQENTAQTSAYIOYEPHLLFSSSESTLHREYHP 956
Db 878 QENTSLYMNENTYIYLSATSDSEEPETLSDVNVNISIPVKEVEGLQFYSSASEYHISIAA 937
QY 957 YGLTP-----VGPRPEKTLRYQNLGCVYVSGLIISALLPVAHGN---YFLSLQV 1007
Db 938 NETVEVEVNSTEDIGNEINIFYLIRKSSFPMBELKISIFPNMTSNQYPVLYPTGLSS- 996
QY 1008 ITNNASCIVQNLTP-----PGPVHPEDELOHTRLNGSTQCQVAVCHLQGLAKTEV 1061
Db 997 -SENANCPHFEDFDSINSKMTTSTDLKRGITLDONCKPATITCNLTS-SDIQV 1054
QY 1062 SVGLRLVYVNEFPRAKFKSLTVSTFELGTEEGSVLQLTASRMSSESLLEVQOT-RPIL 1120

DB 1055 NWSL--ILWKPFTIKYFSSSILMTTNGELRSENAS-LVUSSNOKRELAIOISKDGLPCR 1111
QY 1121 ISLMILIGSVLGLLALLVFLCMLGCPFAHKKIPEEKREB 1163
DB 1112 VPLWVLITLAFAGLLMLLMLLWKIGF---KRLKKMKEX 1151

RESULT 5
ID ITA2 HUMAN STANDARD; PRT; 1181 AA.
AC P17301;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GP1a)
DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
GN ITGA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 30-44.
RC TISSUE=Endothelial cells;
RX MEDLINE=89308879; PubMed=2545729;
RA Takada Y., Hemler M.E.;
RT "The primary structure of the VLA-2/collagen receptor alpha 2 subunit
RT (platelet GPIa): homology to other integrins and the presence of a
RT possible collagen-binding domain.";
RL J. Cell Biol. 109:397-407(1989).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.
RX MEDLINE=98019223; PubMed=9353312;
RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;
RT "Crystal structure of the I domain from integrin alphabeta1.";
RL J. Biol. Chem. 272:28512-28517(1997).
RN [3]
RP VARIANT HPA-5 (BR).
RX MEDLINE=94043762; PubMed=7901236;
RA Santoso S., Kalb R., Walke M., Kiefel V., Mueller-Eckhardt C.,
RA Newman P.J.;
RT "The human platelet alpha2B integrin Br(a) and Brb are associated with a
RT single amino acid polymorphism on glycoprotein Ia (integrin subunit
RT alpha 2).";
RL J. Clin. Invest. 92:2427-2432(1993).
RN [4]
RP VARIANT GLU-534.
RX MEDLINE=20206009; PubMed=10744142;
RA Kroil H., Gardemann A., Fechter A., Haberboesch W., Santoso S.;
RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G
RT gene polymorphism on coronary artery disease and acute myocardial
RT infarction.";
RL Thromb. Haemost. 83:392-396(2000).
CC -I- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN
CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE
CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
CC EXTRACELLULAR MATRIX.
CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
CC ASSOCIATES WITH BETA-1.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INTEGR) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -I- POLYMORPHISM: POSITION 534 IS ASSOCIATED WITH PLATELET-SPECIFIC
CC ALLOANTIGEN HPA-5 (BR). HPA-5A/BR(A) HAS LYS-534 AND HPA-5B/BR(B)
CC HAS GLU-534. HPA-5B IS INVOLVED IN NEONATAL ALLOIMMUNE
CC THROMBOCYTOPENIA (NAT OR NATP). THE K534E POLYMORPHISM MAY PLAY A
CC ROLE IN CORONARY ARTERY DISEASE (CAD).
CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

CC -I- DATABASE: NAME=PROV; NOTE=CD guide CD49b entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49b.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X17033, CA834894.1; -
CC PIR, A33998, A33998.
CC PDB, IAOX; 25-NOV-98.
CC Genew; HGNC:6137; ITGA2.
CC MIM; 192974; -
CC InterPro; IPR000413; Integrin_alpha.
CC InterPro; IPR002035; VWFA.
CC Pfam; PF00092; VWA; 1.
CC Pfam; PF00357; Integrin_A; 1.
CC Pfam; PF01839; FG-GAP; 5.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00191; Int_alpha; 5.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS50234; VWFA; 1.
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC Platelet; Signal; Repeat; Calcium; Magnesium; Polymorphism;
CC 3D-structure.
CC SIGNAL 1 29
CC CHAIN 30 1181
CC DOMAIN 30 1132
CC TRANSMEM 1133 1154
CC DOMAIN 1155 1181
CC REPEAT 45 103
CC REPEAT ? 378
CC DOMAIN 188 378
CC REPEAT 378 433
CC REPEAT 434 486
CC REPEAT 488 549
CC REPEAT 551 610
CC REPEAT 615 667
CC CA_BIND 499 507
CC CA_BIND 563 571
CC CA_BIND 627 635
CC SITE 1157 1161
CC DISULFID 83 92
CC DISULFID 680 737
CC DISULFID 789 795
CC DISULFID 865 876
CC DISULFID 1019 1050
CC DISULFID 1055 1060
CC CARBOHYD 105 105
CC CARBOHYD 112 112
CC CARBOHYD 343 343
CC CARBOHYD 432 433
CC CARBOHYD 460 460
CC CARBOHYD 475 475
CC CARBOHYD 699 699
CC CARBOHYD 1057 1057
CC CARBOHYD 1074 1074
CC CARBOHYD 1081 1081
CC VARIANT 534 534
CC
CC SQ SEQUENCE 1181 AA; 129295 MW; 7E1B7ED968A94070 CRC64;
CC
CC Query Match 28.8%; Score 1756; DB 1; Length 1181;
CC Best Local Similarity 34.3%; Pred. No. 2.1e-118;
CC Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;
QY 11 LPVPLTLL-----CSFPLNDHHRLLFGPPEAEAGYGVLDVHGVGCGQWMLVGAPWD 63
DB 11 LPLLVLTALSGGIINCCIAVNVGLPEAKIFSGPSSEQFGVAVQGFINDKGNMLLVGSWS 70

QY 64 GPSGDRGDYVRCVPVGAHNPACAKGHT--GDYOLGNSHPAVNMHLSMILETDDGDM 122
 DB 71 GFENMGDYKCPV-DLSTATCEKLNLOTSTSI PNTTEKMTMSGLITRMGIGGFL 129
 QY 123 ACAPLMSRACGSSVFSGICARVDASFQOGSIAPTAQRCPYMDVYVILDGNSIYPM 182
 DB 130 TCGPLMAOOCNGNYTTGVGCSIDSPDOLASFSFPAIQPCPSLIDVVCDESNSTYPPND 189
 QY 183 EYQTFRLVYGKFLIDPEQIOVGLVYGESPVHEMSIGDRTEEVVRAKULSREGRE 242
 DB 190 AVNPFLEKFOGIDIGFTKIQVGLIQYANNPRVFNMTYKTEEMIVATSGTSQYGDL 249
 QY 243 TKTAQAIMVACTEFGFOSHGPREAARLVVVTGDSHDEELPALKACEAGRVRYGI 302
 DB 250 TMTFGAIQYARKAYASAGRSATKVMVVTGDSHOSMLKAVIDCNHNIARFPI 309
 QY 303 AVLGHYLRORDPSSFLREIRTIASDPDERFFNVNTEALDIDVALGDRIFGLEGSHA 362
 DB 310 AVLGYNRNALDITKNLIKAIKIASIPTEHYFFNVDEALTEKAGTLEQIFSIGET-V 368
 QY 363 ENSSSFGLEMSOIGFST--HRLKDGLFGVVGAYDMGCVLM--LEGHLLFPRMALDEB 419
 DB 369 QGGDNFQWESQVFSADYSSQNDIIMLGAVGAFMSGTIVQKTSHGHLFP---KQA 423
 QY 420 FPPALO--NHAAYLGYSVSSMLRGGRRLFLSGAPFRHKGVIATOLKKDGAVVAQGL 477
 DB 424 PQDILDRNHSSYLGYSVAA--ISTGSETHVACAPANTGQVLVSVMENGNITVIOAH 482
 QY 478 QGEQIISYFGSELCPDITDRTDGTVDVLAAPMFLGPONKEGRVYV---LVGQSG 532
 DB 483 RGQIISYFGSVYCSVDVDKDITTDVLLVGAAPYMSDLKKEBERVYVLFITKQIILQOHF 542
 QY 533 LTLQGTLOPEPPDARFGFAMGLPDLNDGFAVAVGAPLEGGHOGALVLYHGOSGR 592
 DB 543 --LEG--PEGIENTFGSAIALSDINDGFNDVIVGSELEQNSGAYIYINHGOTIR 597
 QY 593 PHPAORIAAA--SMPLASYFGSVDRGLDLDGDDVDVAVAGQAAILLSRPVHLTP 650
 DB 598 TKYSQKILGSDGAFRSHLQYFGRSLDGYGLNDSTIDVSIAGFGVQVLMGSIADVAI 657
 QY 651 SLEVTQALISVORDCRRRQGEAVCLTALCFQVYTRTEGRMDHQYMFYSLDEWTG 710
 DB 658 EASFPEKITLVKNQAQ-----IILKICFSAKER-PTKONNOVAIYVNTILDA---- 704
 QY 711 ABAAPFGSGORLSRRL-----RLSGNV-----TCQGLAHVLDTSQVLRVALLTV 758
 DB 705 ----DGFSSRYTSRGLFKENNERCLOKNNVNVQAQSCPHIITYQEPSDVVNSLDLRVD 759
 QY 759 FALDNTTKPG--PVLNEGSPITSIOKLVPFSKDCGPNECVTDVLVQNMDIRGSRKAPV 816
 DB 760 ISLEN---FQTSALFAVSETAKVFSIPFHKDCGEGELCISDVLVDVR-QIPAAQOPPI 815
 QY 817 VAGGRKVLVSTLENKRENAVNTSIIIFSRNLHLASLTPORESI---KYEC-AAPSA 872
 DB 816 VANQNRRLTFSSVYLVKKKRESAVNTGIVVFSENLFFASFS---LPVDTEVTCCVAAQ 871
 QY 873 HARLGVGHVPOTGAKTFLFEFSSCSLSQVFGKLTASDSLERGTIOENNAQS 932
 DB 872 KSVACVGVGALRREGQVFTTFINFDNLONLQOASLSFOALSESQENKA--DNLVNIK 929
 QY 933 AYIOYEPHLLFSSSESTLHRYEVHPTLP-----VGPPEFTTLRVONLCYVVSGLI 986
 DB 930 IFLVDABEHLTRSTININFEIISDQNVSIYHSEFVDPKFIPLSKV--TTGSVPVSMAT 988
 QY 987 ISALLPVAHAGNYFLSLQVITNNASCIVONLTP-----GPPVPELOHTNRL 1038
 DB 989 VLIHPIQYTKENKPLMYLITGVOTDKAGDISCANADIMPLKIGOTSSVSFSKSNFHTKEL 1048
 QY 1039 NSGNTQCOVVRCHLQGLAGTEVSGLLALVHNEPRAKFSITVVSFEELGTEBGSLV 1098
 DB 1049 NCRITASCNVTCMLKDVHMKGEYFVAVNTTRIMNGTFASSTFQVQVLTAAABINTYNPEY 1108

QY 1099 OLTEASRMSSELEVVQTRPILISLMI-----LIGSVLGLLALLVFCIM 1145
 DB 1109 -----VIEDNTVTLPMIMKPEKAEVPTGVIIGSIAGILLALLVALIM 1154
 QY 1146 KLGFFPAHK-----KTFEE 1158
 DB 1155 KLGFFKRYEKMTKNPDE 1172
 RESULT 6
 ITA2_BOVIN
 ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
 AC P53710;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GP1a)
 GN (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCB1_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94193647; PubMed=7511592;
 RA Kamata T., Puzon W., Takada Y.;
 RT "Identification of putative ligand binding sites within I domain of
 RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
 RL J. Biol. Chem. 269:9659-9663(1994).
 CC -I- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
 CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE
 CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX.
 CC -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, L25886; AAB59255.1; -.
 CC HSSP: P17301; IAOX.
 DR InterPro: IPR000413; Integrin_alpha.
 DR InterPro: IPR02035; VWF_A.
 DR Pfam: PF00092; vwa; 1.
 DR Pfam: PF00357; Integrin_A; 1.
 DR Pfam: PF01839; FG-GAP; 5.
 DR SMART: SMO0191; Int_alpha; 5.
 DR SMART: SMO0327; VWF_1.
 DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE: PS50234; VWFA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.
 FT NON TER <1 1
 FT SIGNAL 1
 FT CHAIN 19 1870 POTENTIAL.
 FT DOMAIN 19 1121 INTEGRIN ALPHA-2.
 FT TRANSMEM 1122 1143 EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.

```

FT DOMAIN 1144 1170 CYTOPLASMIC (POTENTIAL).
FT REPEAT 34 92 FG-GAP 1.
FT REPEAT ? 92 FG-GAP 2.
FT DOMAIN 177 367 VMEFA.
FT REPEAT ? 92 FG-GAP 3.
FT REPEAT 423 475 FG-GAP 4.
FT REPEAT 477 538 FG-GAP 5.
FT REPEAT 540 599 FG-GAP 6.
FT REPEAT 604 656 FG-GAP 7.
FT CA_BIND 488 496 POTENTIAL.
FT CA_BIND 552 560 POTENTIAL.
FT CA_BIND 616 624 POTENTIAL.
FT SITE 472 474 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1146 1150 GEFKR MOTIF.
FT DISULFID 72 81 BY SIMILARITY.
FT DISULFID 669 726 BY SIMILARITY.
FT DISULFID 778 784 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 1008 1039 BY SIMILARITY.
FT CARBOHYD 1044 1049 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 580 580 G -> V.
FT VARIANT 588 588 R -> K.
FT VARIANT 725 725 R -> S.
SQ SEQUENCE 1170 AA; 128929 MW; EECERIGSF2448BBI CRC64;

```

Query Match 28.7%; Score 1754; DB 1; Length 1170;
 Best Local Similarity 33.9%; Pred. No. 2.8e-118;
 Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

```

QY 9 LFLPLVFLTLG---CSFPLDHEHRLPPGPEAFGYSVLOHVGGGGGMVLVGAHPMG 65
DB 2 LQLVTVFQGLNCCVAVNVGLPKAKIFSGPSSQFQYAVQOPIPKGMWLVGSPMG 61
QY 66 SGDRGDVYRCPCVGAHNAAPCAKGL-GDYOLGNSHPVANNHGLMSLLETDDGGFMAC 124
DB 62 PKNRGADYKCPV-DLSTTTCKNLQSTMSNVTENKTNLSGLTLTRNVGTGGLTC 120
QY 125 APLMSRACGSSVFSGGICARVDASFQPOGSLAPTAQRCPTVMDVIVLDGNSIYPMSEV 184
DB 121 GPLMAGQCGSQYTTTGVCSDDVSPDQLRTSFAPAQVOTCSFIDVVVVCDESISYPMNAV 180
QY 185 QTFLERLVGLFIDEQIOVGLVOYGESPHEMSIGDRTKEVYRAKUNSRREGRTK 244
DB 181 KNFLEKFGQGLDIGTKQWGLIOYANNRPPVFNNTFKSKDEMIKASOTFOYCGDITN 240
QY 245 TQAQITVACBEGFSGHSGRPREARLLVYVTDGSHDEBELPAALACAGAVTRGIAY 304
DB 241 TPKAIQYARDTAYSTAAGRPATKVMVYVTDGSHDSKLAAYVDQCKKNILRFGLAV 300
QY 305 LGHYLRORDPSSFLREIRTTASDPERFFENVTDDEALTDIVDALGRIFGLESHAHN 364
DB 301 LGYLNRNALDTKNLKEIKALIASIPTEHFFNVSDADLEKAKGTIGGIFIEGT-VQG 359
QY 365 ESSFELMSGQICFSTRLLKDG--IIFGMVAYDWGGSVLM-LEGHRLFPFMALEDEFP 421
DB 360 GDNFQWMSQVGFSAEYSPONNIIMLGAYGADWAGTVOAKTPHHLIFS-----KQAFE 414
QY 422 PALQ--NHAAYLYGVSVMILRGGRPLFSGAPRRHGRKATFOLKKDGAIVAVASIQG 479
DB 415 QILQBRNHSSTYIGIVAS--ISTGNSVHFVAGAPRANVTQIYLVISNENGVATVYQSORG 473
QY 480 EOIGSYFGSELCPDLTDRTDGTDLVLLVAPMFLGPNKKEGRVYVYLVGQOSLTLTGTL 539

```

```

DB 474 DQISYFSSVLCADVNDKTTITDVLVGAHPYNDLKEBGRVYLFTTKGILMWHQFLE 533
QY 540 QPEPRDAPFGFAMGALBDLNOGFAVAVAGAPFEDHGQALYHYHGOSVRRBHPAORI 599
DB 534 GPNGLNENKFGSALIALSDIMMAGENDYVCSPLENONGAVIYNGHEGRIRLYSOKI 593
QY 600 AAASMPHA--LSYGRSVYDGLDLGDDLVDVAVAGCAAILSSRPVHLTPSLEVTPQ 657
DB 594 LGSBRASFSHLYQYGRSLDYGDLNGDSITDVSVGAQGVVQVLMQSIAVSVASFPFK 653
QY 658 AIVSVQDRCRRRGQBAVCLTALCFQVTSRTPGRMDHOFYRFTASLDEMTAG-----ARA 713
DB 654 KITLKNKAE-----IKLKLCSAKFR-PTNQNQVAIVYITIDEDQFSSRSVRG 704
QY 714 AFDGSGQLSRRLRLSVGNTVCSQHLNHHVLDTSQDYLAPVLAJTFPLADNTTKG--VVL 771
DB 705 LFKENNEKCLQKTIWISQAO-RCEYIITHIQSPDILSPNLQNMISLEN--PETNPAL 760
QY 772 NEGSPTSIQKLVPFSKDCGPNCEVTDLVLOVNMIDGSRKAPVVRGRRKVLVSTTLE 831
DB 761 EAYSETVAVFSPHKKDGDGVCISDVLNV-QOLPATQOQPIVSNQNRILTFVQLK 819
QY 832 NRKENAVYTSLSIFSRVHLASLTPORESPT--KYEC-AAPSAHARLCSVHPVFQTG 887
DB 820 NKKSAYVTEIVDPFSENLFPASWS---MEVDGTEVYVQYLAASQKSVTCNVGYPALSK 875
QY 888 AKVFLLEFEESGSLISQVFGKLTASDSLERNGTLOENAOVSATYQVEPHLFSSES 947
DB 876 QQVFTTFIDFNLOLQWASISFRALSESEEN--MADNSVNLKLSILYAEIHTST 933
QY 948 TLHRYE-----VHPYGLPVGPGPEKTLRYONLCGYVSGIILSALLPAVANG 997
DB 934 NINFEVSLDGNVSVHVSFEDI---GPKTIFSIKV-TTGSVYVSNASVYIHIPOYTKD 988
QY 998 GNYFLSLQVITNNA---SCIVONLTBPGRPVHP-----EELQHTNRING 1040
DB 989 KNPLMYLTGVATDQAGDISCAE-----INPLKIGQTSSSVSFKEENFRHKEINLC 1039
QY 1041 SNTQGVVRCGLQGLAKTEVSLGLRLVNHNEFFRAKFKSLYVSTRELTGEBSSVQL 1100
DB 1040 RTASCSTNMCWLRDQVKGFEFLVNVSTRIMGTAPASTPQVQVLAAEIDITVNPQIYVI 1099
QY 1101 TEASRWSESLLEVQVTP---ILISLWILISVGLLILALLVFLCMLKGFPAHK----- 1153
DB 1100 EE-----NTVTIPLTIMKHEVVEVPTGVISVAGIULLALVAILMKLGFPRKRYEKM 1155
QY 1154 -KIPEEKREBEKL 1165
DB 1156 AKNPDETDETEL 1168

```

RESULT 7
 ITA2_MOUSE STRAND; PRT; 1178 AA.
 AC 062469; 062163;
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-1998 (Rel. 36, Last sequence update)
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GPIa)
 GN (Collagen receptor) (VLA-2 alpha chain) (CD49b).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC MEDLINE=C57BL/6 X CBA; TISSUE=lung;
 RX Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,
 RA Darganovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
 RA "The mouse VLA-2 homologue supports collagen and laminin adhesion but

RT not virus binding." ;
 RL Cell Adhes. Commun. 2:131-143(1994).
 RN [2]
 RP SEQUENCE OF 450-1178 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=94355691; PubMed=7521231;
 RA Wu J.E., Santoro S.A.;
 RT "Complex patterns of expression suggest extensive roles for the alpha
 RT 2 beta.1 integrin in murine development." ;
 RL Dev. Dyn. 199;292-314(1994).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOZOUS FOR A NULL MUTATION
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INTEGRIN) IS A VMPA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z29987; CA82877.1; -;
 DR EMBL; X75427; CA53178.1; -;
 DR HSSP; P17301; IAOX.
 DR MGI; MGI:96600; Itga2.
 DR InterPro; IPR0000413; Integrin_alpha.
 DR InterPro; IPR002035; VMP_A.
 DR Pfam; PF00092; vwa; 1.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PR00453; VMPADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA_1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VMPA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 DR Platelet; Signal; Repeat; Calcium; Magnesium.
 KW SIGNAL; 1
 FT CHAIN 1 1178 INTEGRIN ALPHA-2.
 FT DOMAIN 27 1129 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1130 1151 POTENTIAL.
 FT DOMAIN 1152 1178 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 42 100 FG-GAP 1.
 FT REPEAT 185 375 FG-GAP 2.
 FT DOMAIN 185 375 VMPA.
 FT REPEAT 431 483 FG-GAP 3.
 FT REPEAT 485 546 FG-GAP 4.
 FT REPEAT 548 607 FG-GAP 5.
 FT REPEAT 612 664 FG-GAP 6.
 FT REPEAT 664 716 FG-GAP 7.
 FT CA_BIND 496 504 POTENTIAL.
 FT CA_BIND 560 568 POTENTIAL.
 FT CA_BIND 624 632 POTENTIAL.
 FT SITE 480 482 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1154 1158 GPEFR MOTIF.
 FT DISULFID 677 89 BY SIMILARITY.
 FT DISULFID 786 794 BY SIMILARITY.
 FT DISULFID 862 873 BY SIMILARITY.
 FT DISULFID 1016 1047 BY SIMILARITY.

FT DISULFID 1052 1057 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1054 1054 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;
 Query Match 28.6%; Score 1744; DB 1; Length 1178;
 Best Local Similarity 35.3%; Pred. No. 1.5e-117;
 Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;
 9 LFLPLVLT-GL--CSPNLDHHRRLPPRPEAFEGSYVLOHVGCGRRMLVCAFWGCP 65
 10 LLLQLMLLVQGLINCLAVNGLPGAKIFGSGSSSEQFGYSVOQLTNPQGNMLLVGSPWGSF 69
 66 SGDRKGVYRCPCVGAHNAPOAKGHL-GDYOLGNSHFAVMHIGMSLLETDGCGFMAC 124
 70 PENRGDVYKCPV-DLPATCEKLNLSASISNTEIKTNVSLGLTLRNPDTGGLTC 128
 125 APLWRACGSSVFSGICARVDASFPOGSLAPTQRCPTVMDVYIVLDSNSIYPSSEV 184
 129 GLPLMHQGNQYVATGICSDVSPDFQFLRFSRPAQACPSLVDDVYVCDENSITPWEAV 188
 185 QTFRLRLVGLFLPDEQIOVGLVOGESPVHEWSIGDFRYKEEVVRAKNLSRREGRET 244
 189 KNFLVFPVTGLDIGRKTQVALIOYANPRRIIFINLDETEKEDMWQATSETRHOGGLTN 248
 245 TQAALMVACTGSEFSGSHGSRPEARLLVYVNDGSHDEBELPALKACEAGVTRFYIAV 304
 249 TFRALFARDVAYVSGRPGATVWVVDGSHSDSKKTYVQQCNDEILRFGLAV 308
 305 LGHYLRORDPSSFLREIRTIASDPDEFNFVNTDEALTDIVDALGRIFGLESHAEN 364
 309 LGYLNRNALDTPKMLIKEIKALIASPTERYFFNVADDEALLKAGTLGQIFSIEST-VQG 367
 365 ESSFGLENSQIGFSTHRL-KDGILFGVNGAYDNGSVLMLEGGHR--LFFPRNMLEDEF 420
 368 GDNFMENAVQVGSADYAPONDIMLGAVGAFDWSGLTV-DETSKRPVIFP---KQAF 421
 421 PPALO--NHAAYVIGVSVMILRGGRFLFEGAPFRFRGKVIYFOLKKDGAVARVAGSLQ 478
 422 DOVLDRNHSSFLGVSVAISTEDGVH-FVAGAPRANVTGQVLYSVKKGQVNYIQSHR 480
 479 GEOLGVSFGESELCPLDTRDGTDDVLLVAAPFLGPONKETGRVYVVLVGOQSLTLTQGT 538
 481 GQIGISYFSGVLCSDVDVDKITTDLVVGAPFTYNNMDLKEGKQVLPFITRIGILNQHF 540
 539 LQPEPPDARFGPMGALPDLNODGFADVAVGAPLEDHOGALYLYHGTQSCVRPHRQR 598
 541 EGPEGTGARFGSALALASDINMGFNDVIGSPVENNSGAVYVYNHOGTIRKYSQK 600
 599 IAAA--SMPHALSYRGSRGRLDGDPLDVVAAGQAAILLSRPIVHLTPSLVETP 656
 601 ILGSGATRRHLQFGRSLDGYGDLNGSDITDVSIGALGOYIQLMSQSLADVAIALTLTP 660
 657 QAISSVQDRCRRRGGEAVCLTALCFQVTSRTPGRMHDQFYWRFTASLDEWTAGARAAD 716
 661 DKITILNNDK-----ITLKLCFRPFPPAQQ--NNQVALIFNMTLDDAGHSRVS 711
 717 GSGQRLSPRLR--LSGVNV--TCEQLHFLVLDTSYLRPVALTVTFALDNTTKPC--PVL 771
 712 GVFRNSERFLQKMNVAVEQKSEHHISIQKPSDVVNPDLRDVDSILEN--PETSAL 768
 772 NEGSTSIQKLVPSKDCGPDNECTDVLQVMDIRSRKAPFVVRGRGRKVLVSTLE 831
 769 EAYSTVAVFSIPFYKCGSDGICISDLILV-QQLPAIQOSFVSQNRKLTFVSILK 827
 832 NRKENAVNTSLSIFSRMLASLTPQRESPI---KYECAPSAHARL-CGVGHPVQTC 887

```

Db      828 NRGESAYNTVVALEFENLEFASFS-----MPVDGTVTEVGVSQKSVTCVGYPALKSE 883
Oy      888 AKVTFLEEFSSCSLSQVFGKLTASPSLENGTLOENTQTSAYIQEPLHFSSSS 947
      884 QVTFITINDENQONQONQAINFOAFSSQETNKA--DNSVSLITPLIYDELHTRST 941
Oy      948 TLHREVPHPYGLP-----VGPPEFKTLRQVNYGCVVSGSLISALLPAVHAGNYF 1001
      942 NINFEYISDENAPSVIKSVEDIGPKFISLKV-TAGSAPVSMALVTIHIPQYTEKNPL 1000
Oy      1002 LLSQVITNNA---SCL--VQNLTEP---PGPVHHEELQHTNRLNGSTQCQVVRCHG 1053
      1001 LVLGTQTDAGDISCTAEINPKLPHTPASVSFKXNENRHTKELDCTRITSCSNITCWLK 1060
Oy      1054 QAKGTEVSVGLRLVHNEFFRPAKSLTVSTFELGTEEGSVLOLTFASRWSESLLEV 1113
      1061 DLHMAEYFINVTRWRNRTFAASTQYQVQLDPAEIDHPNQ-LFVIEENAVTIPLMIM 1119
Oy      1114 VQTPRLISLWILIGSVLGLLLALLVFCWLKGF--AHKKI---PEE 1158
      1120 KPEKAEVPTGYIGIIGILLMTAGLWKLGFGRKQYKKMGQNPDE 1169

```

RESULT 8

```

ID      ITAD_HUMAN      STANDARD;      PRT; 1162 AA.
AC      Q13349; Q15575; Q15576;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-D precursor (Leukointegrin alpha D) (CD11d) (ADB2).
GN      ITGAD.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
      (1)
RP      SEQUENCE FROM N.A.
RC      TISSUE=Spleen;
RX      MEDLINE=96111956; PubMed=8777714;
      Van der Vlieten M., Le Trong H., Wood C.L., Moore P.F., St John T.,
      Staunton D.E., Gallatin W.M.;
      "A novel leukointegrin, alpha d beta 2, binds preferentially to ICAM-
      3.";
      Immunity 3:683-690(1995).
RN      [2]
RP      SEQUENCE OF 1-235 FROM N.A.
RX      MEDLINE=20187620; PubMed=10722744;
      Nott J.D., Johnson A.K., Dillon J.D.;
      "Structural and functional characterization of the leukocyte integrin
      gene CD11d. Essential role of Spl and Sp3.";
      J. Biol. Chem. 275:8959-8963(2000).
RN      [3]
RP      SEQUENCE OF 467-537; 571-602; 633-666; 788-834 AND 910-1125 FROM N.A.
RX      MEDLINE=96257236; PubMed=8666289;
      Wong P.A., Davis B.M., LeBeau M., Springer T.A.;
      "Cloning and chromosomal localization of a novel gene-encoding a human
      beta 2-integrin alpha subunit.";
      Gene 171:291-294(1996).
RN      [4]
RP      INTERACTION WITH VCAM1.
RX      MEDLINE=99059842; PubMed=9841932;
      Grayson M.H., Van der Vlieten M., Sterbinsky S.A., Michael Gallatin W.,
      Hoffman P.A., Staunton D.E., Bochner B.S.;
      "alpha-defensin integrin is expressed on human eosinophils and functions
      as an alternative ligand for vascular cell adhesion molecule 1
      (VCAM-1).";
      J. Exp. Med. 188:2187-2191(1998).
RN      [5]
RP      INTERACTION WITH VCAM1.
RX      MEDLINE=99370002; PubMed=10438935;
      Van der Vlieten M., Crowe D.T., Hoekstra D., Vazeux R., Hoffman P.A.,

```

```

RA      Grayson M.H., Bochner B.S., Gallatin W.M., Staunton D.E.;
RT      "The leukocyte integrin alpha D beta 2 binds VCAM-1: evidence for a
      binding interface between integrin alpha domain and VCAM-1.";
      J. Immunol. 163:1984-1990(1999).
RL      [1]
CC      FUNCTION: INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM1 AND
      VCAM1. MAY PLAY A ROLE IN THE ATHEROSCLEROTIC PROCESS SUCH AS
      CLEARING LIPOPROTEINS FROM PLAQUES AND IN PHAGOCYTOSIS OF BLOOD-
      BORNE PATHOGENS, PARTICULATE MATTER, AND SENESCENT ERYTHROCYTES
      FROM THE BLOOD.
CC      SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-D
      ASSOCIATES WITH BETA-2.
CC      TISSUE SPECIFICITY: EXPRESSED MODERATELY ON MELANOCYTIC CELL
      LINES AND SUBSETS OF PERIPHERAL BLOOD LEUKOCYTES AND STRONGLY ON
      TISSUE-SPECIALIZED CELLS, INCLUDING MACROPHAGES FOAM CELLS WITHIN
      ATHEROSCLEROTIC PLAQUES, AND ON SPLENIC RED PULP MACROPHAGES.
CC      DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
      WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      SIMILARITY: CONTAINS 1 WFPA DOMAIN.
CC      SIMILARITY: CONTAINS 7 FG-GAP-REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (see http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).
      -----
DR      EMBL; U37028; AAB38547.1; -
DR      EMBL; U40274; AAB60634.1; -
DR      EMBL; U40275; AAB60635.1; -
DR      EMBL; U40276; AAB60636.1; -
DR      EMBL; U40277; AAB60637.1; -
DR      EMBL; U40279; AAB60638.1; -
DR      EMBL; U40278; AAB60638.1; JOINED.
DR      EMBL; AF187881; AAF62875.1; -
DR      HSSP; P11215; IABX.
DR      Genew; HGNC:6146; ITGAD.
DR      MIM; 602453; -
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; WFPA_A.
DR      Pfam; PF00092; vwa; 1.
DR      Pfam; PF00357; Integrin_A; 1.
DR      Pfam; PF01839; FG-GAP; 5.
DR      PRINTS; PR01185; INTEGRINA.
DR      PRINTS; PR00453; WFPADOMAIN.
DR      SMART; SM00191; Int_alpha; 4.
DR      SMART; SM00327; VWA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; WFPAR; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Signal; Repeat; Calcium;
KW      Magnesium.
FT      SIGNAL          1      17
FT      CHAIN           18     1162
FT      DOMAIN         1101    1124
FT      TRANSMEM      1125    1162
FT      DOMAIN         1125    1162
FT      REPEAT         32     85
FT      REPEAT         150    332
FT      REPEAT         350    400
FT      REPEAT         401    452
FT      REPEAT         454    516
FT      REPEAT         518    576
FT      REPEAT         581    633
FT      CA_BIND        465    473
FT      CA_BIND        530    538
FT      CA_BIND        593    601
FT      SITE           1127    1131
FT      DISULFID       67      74
FT      POTENTIAL.
FT      INTEGRIN ALPHA-D.
FT      EXTRACELLULAR (POTENTIAL).
FT      POTENTIAL.
FT      CYTOPLASMIC (POTENTIAL).
FT      FG-GAP 1.
FT      FG-GAP 2.
FT      WFPAR.
FT      FG-GAP 3.
FT      FG-GAP 4.
FT      FG-GAP 5.
FT      FG-GAP 6.
FT      FG-GAP 7.
FT      POTENTIAL.
FT      POTENTIAL.
FT      POTENTIAL.
FT      GPFR MOTIF.
FT      BY SIMILARITY.

```

FT	DISULFID	106	124	BY SIMILARITY.
FT	DISULFID	655	710	BY SIMILARITY.
FT	DISULFID	769	775	BY SIMILARITY.
FT	DISULFID	846	861	BY SIMILARITY.
FT	DISULFID	994	1018	BY SIMILARITY.
FT	DISULFID	1023	1028	BY SIMILARITY.
FT	CARBOHYD	59	59	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	87	87	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	99	99	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	331	331	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	733	733	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	873	873	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	957	957	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1046	1046	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	500	500	MISSING (IN REF. 2).
FT	CONFLICT	515	518	GHPW -> ATP (IN REF. 2).
FT	CONFLICT	825	825	L -> V (IN REF. 2).
FT	CONFLICT	984	984	V -> A (IN REF. 2).
SQ	SEQUENCE	1162 AA;	126885 MM;	F296A1A35455D7DD CRC64;

[4]
 X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 153-335.
 MEDLINE=96398682; PubMed=805579;
 Qu A., Leahy D.J.;
 "The role of the divalent cation in the structure of the I domain
 from the CD14/CD18 integrin";
 Structure 4:931-942 (1996).
 [5]
 X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 153-334.
 MEDLINE=99425288; PubMed=10493852;
 Kallen J., Welzenbach K., Ramage P., Geyl D., Krawacki R., Legge G.,
 Cottens S., Weitz-Schmidt G., Hommel U.;
 "Structural basis for LFA-1 inhibition upon lovastatin binding to the
 CD14 I-domain";
 J. Mol. Biol. 292:1-9 (1999).
 - I- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 ICAM3 AND ICAM4. IT IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA,
 INCLUDING LEUCOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL
 MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES
 AND MONOCYTES.
 - I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L
 ASSOCIATES WITH BETA-2.
 - I- SUBCELLULAR LOCATION: Type I membrane protein.
 - I- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 produced by alternative splicing.
 - I- TISSUE SPECIFICITY: LEUKOCYTES.
 - I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMFA DOMAIN. INTEGRINS
 WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 - I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 - I- SIMILARITY: CONTAINS 1 VMFA DOMAIN.
 - I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 - I- DATABASE: NAME=PROW; NOTE=CD guide CD14 entry;
 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd14.htm".

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 DR EMBL, Y00796; CAA68747.1; -
 DR EMBL, AC002310; AAC31672.1; -
 DR PIR, S03308; S03308.
 DR PDB, 1LEA; 29-JAN-96.
 DR PDB, 1ZON; 07-DEC-96.
 DR PDB, 1ZOO; 07-DEC-96.
 DR PDB, 1ZOP; 07-DEC-96.
 DR PDB, 1COP; 07-AUG-00.
 DR Genew; HGNC:6148; ITGAL.
 DR MIM; 153370; -
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VMF_A.
 DR Pfam; PF00092; vma; 1.
 DR Pfam; PF00357; integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VMFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; vma; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS00234; VMFA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; 3D-structure; Magnesium; Calcium; Repeat;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1170 INTEGRIN_ALPHA-L.
 FT DOMAIN 26 1088 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1089 1112 POTENTIAL.
 FT DOMAIN 1113 1170 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 42 91 FG-GAP 1.
 FT REPEAT 92 149 FG-GAP 2.

FT	DOMAIN	170	349	VMFA.
FT	REPEAT	?	?	FG-GAP 3.
FT	REPEAT	401	455	FG-GAP 4.
FT	REPEAT	457	516	FG-GAP 5.
FT	REPEAT	518	575	FG-GAP 6.
FT	REPEAT	578	630	FG-GAP 7.
FT	CA_BIND	468	476	POTENTIAL.
FT	CA_BIND	530	538	POTENTIAL.
FT	CA_BIND	590	598	POTENTIAL.
FT	SITE	1115	1119	GEFGR MOTIF.
FT	DISULFID	73	80	BY SIMILARITY.
FT	DISULFID	111	129	BY SIMILARITY.
FT	DISULFID	653	707	BY SIMILARITY.
FT	DISULFID	771	777	BY SIMILARITY.
FT	DISULFID	845	861	BY SIMILARITY.
FT	DISULFID	998	1013	BY SIMILARITY.
FT	DISULFID	1021	1052	BY SIMILARITY.
FT	CARBOHYD	65	65	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	726	726	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	730	730	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	862	862	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	885	885	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1060	1060	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	954	954	Q-> QGVGLVETQTSQKICRPFAGDEHVAQGEGLPC PMGSEAFRDINRAGPCR (IN ISOFORM 2). R-> W (IN REF. 1 AND 2). Y-> I (IN REF. 2).
FT	CONFLICT	214	214	
FT	CONFLICT	660	660	
FT	SEQUENCE	1170 AA;	128819 MW;	39A7AF2EF286FC0 CRC64;

Query Match 18.1%; Score 1105; DB 1; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 1.9e-71;
 Matches 366; Conservative 198; Mismatches 465; Indels 216; Gaps 55;

QY	11	LPVLVLTGL-----CSPNLDHHRPLFPGPPEA--EFGSYVLQVHGGQRMVLGAWMD 63
DB	9	MAMALLSGFFPPAPASSTNLDVGRARSF-SPPARRHGKRVLQ-VGNG--VVGAGGE 63
QY	64	GPSGDRRDVYRCIPGAGAHNAPCAKGLGDVQIGNSSHPAVNMILGMSLLETDDGGFMA 123
DB	64	GNS---TSLVQCQCGSTGHCLPYT-----LRGSNY--TSXYLGMTLATDPTDGSILA 110
QY	124	CAPLWSRACGSSVFSSGICARVDASFQ-PQGSIAFTAGRC-PTYWDVVIYLDGNSNIYP- 180
DB	111	CDPGLSRFCQDQNTVYLSGLCYLFRQNLQGMILQGRPGQECIKGNVDLVFLFDGWSLQPD 170
QY	181	MSEVQTFLRLRYVGLFLFDPEQIQVGLVQYGESPVHMSLDF---RKKEVVRARAKULS 236
DB	171	EFQKILDFMKVVMKUL--SNTSYQPAAVQFSTSYKTEFDSYVVRKPDALDKVVKML 228
QY	237	RREGRETKTAQAINVACTEGFSQSHGRPEARLIVVTDSESHGEELPALAKACEAGR 296
DB	229	LL-----TTFGAINVVALEVRREELGARDPATKVIILITDEATDSGNIDAAD----- 278
QY	297	VTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFPNVNTDEALTDIYDALGDRIFG 356
DB	279	IIRYIIIGIKHFQTESOET-----LHKFASKPASEEFKILDTFEKLKDLFTELQKIYV 333
QY	357	LEGSHAENSSFGLEMSQIGFSTHKLKGLIFGNGVAGADMGGSVLMLEGHRLLFPPMAL 416
DB	334	IEGTSKQDTSFNMELSSGSIADISRGHAAVGAAGAKOMAGFLDLDAADQ----- 385
QY	417	EDDEF-----PPALQNHAAVLGYSVSSMLRGGRRFLTSGAPRRHKGKVIATQKKDGA-V 471
DB	386	DDTFIGNPLRPPEVAGVGLGYTWLPSRQKTSLSASGAPRYQHGKRVLLQDEPGGSHW 445
QY	472	RYAQSLQGEQIGSYFGSELCPLDTRDGTVDVLLVAADMPFLGPONKGTGRVYV--LVG 528

Db 446 SQVOTIHGTOIGSYFGGELGVVDVDDGETELLIGAPLPGYEQ--RGGRVFYQRRQIG 503
 QY 529 QQSLLTLOGLOPEPQDARFGFAMGALPDLNODGFADVAVGAPLEDHOGALYLHGNG 588
 Db 504 FEVSSELQ--DGYPL-GRFGAITLALDINDGADVAVGAPLE--QGAAYITNGH 558
 QY 589 SGVRPAPORIAAAMPALSYFGRSVGDLDDGDLVDVAVGAOGAAILSSRPVHL 648
 Db 559 GGISPPSPORIESTQYLSGIQWFGRSIHGVKDEGGGLADVAVGASQMTVLSRRVDM 618
 QY 649 TBSLEVTPOAISVQDC---RRGQEAVALTALCOVSTSTPGRMWHQFMR----- 699
 Db 619 VTLMSFSPAEIPHEVECSYSTSNKKEGVNIT--ICFOIKSLYP-----QFQGRVLVNL 671
 QY 700 -FTASLDENTAGARAFDSCGQSLSPRLSLVSGVNTCEGLHPIV-LDTSDYAPVALTV 757
 Db 672 TYTLQJDGHRTRRGIFPGGRHEL--RRNIATVTSSTCDTFSHFVPCVODLISPINVS 729
 QY 758 TFAL---DNTTKGPVILNESPSTIOKL-----VPSKDCGPDNECVTDVLQVNA-MD 806
 Db 730 NFSLWEEBEGTPDQRAQKDIPIRLPSLHSETWEIPFENKCGEDKKCANLKVSPSPAR 789
 QY 807 IRGSRAPFVYVGRKRVYVSTLLENKKNAYNTSIIIFSRVHLASLTPQK-ESPITY 865
 Db 790 SRALRLTARA---SLSYELSLNLEBDAYVQDLHPGSLSPKXVEMLKPHQIPV 843
 QY 866 ECAPAPAHARL-----CSVGHVFPOTGAKVTFLEFEPSCSLLSQVFGKLTASS--DS 917
 Db 844 SCEELPEBSRLSRALSCNVSPITFAGHSVA-----LQMMENTLVNSSWGS 891
 QY 918 LERNGT-----LQNTAQTSAVIOYEPHLF--SSESTLHRYEVHAPYGLTPVPG 965
 Db 892 VELHANVTNNEDSDLEDNASTIIPILYINILQODEDSTLY-----VSFTKGP- 944
 QY 966 PEKKTILRQNLCTCYVVSGLIITALLPAYAHGNYLSLSQVTTNASCIVQNLTPP-- 1023
 Db 945 -----KIHQVKMYQVRIOPSI-HDN-----IPTLRA---VVGVPDPSE 981
 QY 1024 GP-----VHPELOHTNRKLSNTOC---QVVRCHLQALKGTEVSGVLRL 1067
 Db 982 GPTTHQMSVQMEPPVCHYEDLE---RLPPAPPCPLPGALFPCPVFROELVQVGTLE 1038
 QY 1068 LV---HNEFFRAKRSKLTVST--FELGTEGSAVLQUTLEASRWSESLLEVQTRPILI 1121
 Db 1039 LVGEIRASSWFSLCSSISISIFNSKHFHLYGNASIAQVY-----MKVDVYKROMLY 1091
 QY 1122 SLMLIGSVLGGILLALLVFCIMKIGFPAHKKIPEEKREKYLE 1166
 Db 1092 -LYVLISG--IGGLLLLIPIVLKYKGF-----KRLKEKME 1126

RN [2]
 RP SEQUENCE OF 24-42.
 RX MEDLINE=85188276; PubMed=3887182;
 RA Springer T.A., Teplow D.B., Dreyer W.J.;
 RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion
 RL glycoproteins and unexpected relation to leukocyte interliron";
 CC Nature 314:540-542 (1985).
 CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 CC ICAM3 AND ICAM4. IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA
 CC INCLUDING LEUKOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL
 CC MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES
 CC AND MONOCYTES. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-L
 CC SUBUNIT GENE DEMONSTRATE IMPAIRED TUMOR REJECTION AND IMPAIRED
 CC LEUCOCYTES RECRUITMENT.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L
 CC ASSOCIATES WITH BETA-2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: LEUKOCYTES.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTRAINS 7 FG-GAP REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M60778; AAA9426.1; -
 CC HSB: P20701; ILPA.
 DR MGD; MGI:96606; Itgal.
 DR InterPro; IPR00413; Integrin_alpha.
 DR InterPro; IPR02035; VWF_A.
 DR Pfam; PF00092; vwa; 1.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Magnesium; Calcium;
 KW Repeat.
 FT CHAIN 1 23
 FT SIGNAL 1 23
 FT DOMAIN 24 1163
 FT TRANSMEM 24 1084
 FT DOMAIN 1085 1108
 FT DOMAIN 1109 1163
 FT REPEAT 39 88
 FT REPEAT 88 148
 FT DOMAIN 148 334
 FT REPEAT 334 454
 FT REPEAT 454 514
 FT REPEAT 514 576
 FT REPEAT 576 628
 FT CA_BIND 628 474
 FT CA_BIND 474 536
 FT CA_BIND 536 596
 FT SITE 596 1111
 FT DISULFID 1111 77
 FT DISULFID 77 108
 FT DISULFID 108 126
 FT DISULFID 126 147
 FT DISULFID 147 199
 FT DISULFID 199 705
 FT DISULFID 705 773
 FT DISULFID 773 840
 FT DISULFID 840 1009
 FT DISULFID 1009 993
 FT DISULFID 993 1009

FT DISULFID 1017 1048 BY SIMILARITY.
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 668 668 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 724 724 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 776 776 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 890 890 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 927 927 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 1056 1056 N-LINKED (GLCNAC. . .) (POTENTIAL)
 SQ SEQUENCE 1163 AA; 128343 MW; A7A3078489BE232F CRC64;
 Query Match 18.0%; Score 1097; DB 1; Length 1163;
 Best Local Similarity 28.4%; Pred. No. 7.2e-71;
 Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;
 12 PLVFLTGL-----CSFNLDEHHRLFFRRPRAEKGYSTLQHVGGQRMVLGAPMDGS 66
 8 PRLLLGLQLFAKAMSYNLDTRPTOSFLAAGRHFQYQLIEDG---VVVGAPGE--- 60
 67 GDRSGDYVRCPVGAHANAPCAKAGHLADYQLGNSSHPAVMHGLMSLLETDDGGFMACAP 126
 61 GNTGGVHICRTSSERCQVS-----LHGSNH--TSKYGMTLATDAAGSLACDP 110
 127 LMSRAQSSVFSGGICARVDASFQDQSLAPTAORCPY-----MDVVIILGDSNSI- 178
 111 GLSRCTDQNTYLSGLC-----YLFPQSLGCPMLQNPFAVQCMKGVLDLVPFDQSQSLD 165
 179 -YPMSEVOFLRLRYVKLFIDEBOIQVGVGESVHEHMSLSDPRTKEVYRAKNLSR 237
 166 RQDFEKLTFEMKDVNKL--SNTSYQFAAVFSTDRCTFTFLDY----VKQNNPQV 217
 238 REGRE-----TKTAQAIWVACTEGFSGHGRPEAKRLVVVVTGESHDELPALAKA 291
 218 LIGSVQPMFLNTRAIYVVAHVKESSGARPDATKVLVITTDSEADKNISA---- 273
 292 CEAGVTRYGIAVLGHYLRQRDPSSFLREIRTIASDPERFEFNVTDAAITDIDALG 351
 274 --AHDIRYIIIGIKHFVSVOKO----KTLHIFASEPEVEFKLIDPEKIKDLFTDQ 326
 353 DRIFGEGHAENESSFGLEMQIGSTRKLDGILFGVAVYDMGGSVLMLEGHRLFP 411
 327 RRIYAIEGTNRQDLISFNNELSSSGISADLSKHAVVAGAKDMAAGFLDLR----- 379
 412 PMALED-----EPPALQNHAAVLGYVSSMLRGRLPLSGARFRRHGRKVLAF 463
 380 -----EDLOGATFVGOEPLTSDVGRGYLYTVAWMTSRSLPDLAAGARYVHGVGLVF 434
 464 QV-KKDGAVVAQSLQEOIGSYFSGELCPDLTDRDITDVLVLAAPMLGPONKETGRV 522
 435 QAPRAGGRNMQOKIEGTQIGSYFSGELCSVDLDQGEAKILLIGARLPFGSQ--RGGRV 492
 523 YVYLVGQGSLLTLOGLTQEPERPD-ARFGFAMGALPDLNQDGFADYAVAGAPLEDHOGAL 581
 493 FTY-QRROSLFPMVSELQDPEVPLGRFGAATLALDINGDLITDVAAGAPLEE--QGAV 549
 582 YVYHVGQSVRPHAPORIAASWPHALSYFGRSVSDGRLLDQGDLDVAVVAGQAAILIS 641
 550 YFNGKPGGLSPOPSRIGQAOVFGIRWFGRSIHCVKDLGGRLADLVVAGGGRVVLIS 609
 642 SRPIVHLTSLVLTPOAISVVOQDC-----RRGOEAVCTLAALCFQVTSRTFGRMDHOF 697
 610 SRPVVDVVELSFSPEIRIVHEVSGYSAREQKHGVKKA--CFRIKVLTP-----QOQ 662
 698 MR-----FTASIDETAGARAAFGDSGGRSLFRRLRLSVGNV-----TCEQLHNH-V 744
 663 GLRLANLSTYLOLDGHRMRSRGLFPDGSHEL-----GNTSITPDKSCUDPFHFHPI 714

QY 745 DTSYLREVALVTVPALDNTTKPGPVINEGSP-----SIQKLVPSKSD 788
 DB 745 CIODLISPINSLNLSL-----LEEGTPROCKGRAMOPIRLPSIHTVTKPIPEKN 766
 QY 789 CGPDNECTDLVQVNMIRGSRKAPRVVGRGRKVLVSTLLENKEMAYNTSLISISR 848
 DB 767 CGEDKCEANLTL-----SPARGPLRLMSS-ASLAEVTLNSNGEDAVWVRDLDP 820
 QY 849 NLHLASLTP-QRESPIKVECAAPSAHARL-----CSGVHPVPOTGAKVTFLEPEFSCS 901
 DB 821 GLSPFKVEMLOPHSMRPMVSCBELTEGSSLTKTKCNVSPIFKAGQVS----- 870
 QY 902 SLISGVFGLTASS--DSLERNGLT-----QENTAQTSAYIQYEPHLLFSSSESTL 949
 DB 871 -LQWFMFTLNSGWEDFVELNGVHCENENSSIQEDNSAATHIPVLVPVILTEQGN- 927
 QY 950 HRYEVHPIGTLPVGGPEPFKTKLKYONL-----GCYVSGLLIISALLPVAHGNIFL 1002
 DB 928 -----STLYISPTPGPKTQOVQVYQVRIOPSADYHNPPTLEALVGVPRHSEDLI 979
 QY 1003 SLISQVITNN--ASCIYONLTPRPGPVHPEELQHTNRNGSNTQOVY-RCHLQGLAKGT 1059
 DB 960 YTVMSVQTDPLVTCSEDLKRPSSAEOP-----CLPGVQFRCPVFRWEILIQVTGT 1032
 QY 1060 EVSVGLRLRVHNEPFRRAKFKS-LTVVSTFELGTEEGSVLQLTENSFWS-LEBV 1114
 DB 1033 V-----ELSKETKASTLSLGSLSVFSNFKHFL-VGSKASEAVLVKVDLI 1080
 QY 1115 QTRPILISLMLIGSVLGLLLALVFLCKMGKGFPAHKKIPBEKREKLE 1166
 DB 1081 HEKEML-HVUYLSG--IGLVLFLFLALYKVGFF-----KRNLEKME 1122

RESULT 11
 ID ITAM HUMAN STANDARD; PRT; 1152 AA.
 AC P11215;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (Leukocyte adhesion receptor MO1)
 DE (Neutrophil adherence receptor).
 GN ITGAM OR CR3A OR CD11B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=86315033; Pubmed=2457584;
 RA Corbi A.L., Kishimoto T.K., Miller L.J., Springer T.A.;
 RT "The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b) alpha subunit. Cloning, primary structure, and relation to the integrins, von Willebrand factor and factor B.";
 RL J. Biol. Chem. 263:12403-12411 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86190151; Pubmed=2833753;
 RA Arnout M.A., Remold-O'Donnell E., Pierce M.W., Harris P., Tenen D.G.;
 RT "Molecular cloning of the alpha subunit of human and guinea pig leukocyte adhesion glycoprotein MO1: chromosomal localization and homology to the alpha subunit of integrins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:2776-2780 (1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88257215; Pubmed=2454931;
 RA Arnout M.A., Gupta S.K., Pierce M.W., Tenen D.G.;
 RT "Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor MO1 (complement receptor type 3).";
 RL J. Cell Biol. 106:2153-2158 (1988).
 RN [4]

RP SEQUENCE FROM N.A.
 RA MEDLINE=93123748; PubMed=8419480;
 RX Fleming J.C., Pahl H.L., Gonzalez D.A., Smith T.F., Tenen D.G.;
 RT "Structural analysis of the CD11b gene and phylogenetic analysis of
 RT the alpha-integrin gene family demonstrate remarkable conservation of
 RT genomic organization and suggest early diversification during
 RT evolution.";
 RL J. Immunol. 150:480-490(1993).
 RN [15]
 RP SEQUENCE OF 9-1153 FROM N.A.
 RX MEDLINE=89098893; PubMed=2563162;
 RA Hickey D.D., Hickey M.J., Ozols J., Baker D.M., Back A.L.,
 RA Roth G.J.;
 RT "cDNA sequence for the alpha M subunit of the human neutrophil
 RT adherence receptor indicates homology to integrin alpha subunits.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:257-261(1989).
 RN [6]
 RP SEQUENCE OF 1-9 FROM N.A.
 RX MEDLINE=92073318; PubMed=1683702;
 RA Shelley C.S., Arnout M.A.;
 RT "The promoter of the CD11b gene directs myeloid-specific and
 RT developmentally regulated expression";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10525-10529(1991).
 RN [7]
 RP SEQUENCE OF 1-9 FROM N.A.
 RX TISSUE=Blood;
 RA MEDLINE=9214986; PubMed=1346576;
 RA Pahl H.L., Rosmarin A.G., Tenen D.G.;
 RT "Characterization of the myeloid-specific CD11b promoter.";
 RL Blood 79:865-870(1992).
 RN [8]
 RP SEQUENCE OF 17-31.
 RX MEDLINE=87076671; PubMed=3539202;
 RA Pierce M.W., Remold-O'Donnell E., Todd R.F. III, Arnout M.A.;
 RT "N-terminal sequence of human leukocyte glycoprotein Mol. 1;
 RT conservation across species and homology to placental Iib/Iiia.";
 RL Biochim. Biophys. Acta 874:368-371(1986).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 148-331.
 RX MEDLINE=95171458; PubMed=7867070;
 RA Lee J.O., Rieu P., Arnout M.A., Liddington R.;
 RT "Crystal structure of the A domain from the alpha subunit of integrin
 RT CR3 (CD11b/CD18)".
 RL Cell 80:631-638(1995).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 148-334.
 RX MEDLINE=96363671; PubMed=8747460;
 RA Lee J.O., Bankston L.A., Arnout M.A., Liddington R.C.;
 RT "Two conformations of the integrin A-domain (I-domain): a pathway for
 RT activation?";
 RL Structure 3:1333-1340(1995).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 148-337.
 RX MEDLINE=98362595; PubMed=9687375;
 RA Baldwin E.T., Sawyer R.W., Bryant G.L. Jr., Curry K.A.,
 RA Fairbanks M.B., Finzel B.C., Garlick R.L., Heinrichson R.L.,
 RA Horton N.C., Kelley L.L., Mildner A.M., Moon J.B., Mott J.E.,
 RA Mutchler V.T., Tomich C.S., Watenpaugh K.D., Wiley V.H.;
 RT "Cation binding to the integrin CD11b I domain and activation model
 RT assessment.";
 RL Structure 6:923-935(1998).
 RN [12]
 RP 3D-STRUCTURE MODELING OF 17-616.
 RX MEDLINE=98226734; PubMed=9560195;
 RA Oxvig C., Springer T.A.;
 RT "Experimental support for a beta-propeller domain in integrin alpha-
 RT subunits and a calcium binding site on its lower surface.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:4870-4875(1998).
 CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS
 CC ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES
 CC AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES.
 CC IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF
 CC THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D

CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR
 CC FIRINGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES
 CC OF FIRINGEN GAMMA CHAIN.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-M
 CC ASSOCIATES WITH BETA-2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
 CC GRANULOCYTES.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFPA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11b entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd11b.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL, J03925; AAA59544.1; -;
 CC EMBL, M18044; AAA59491.1; -;
 CC EMBL, J04145; AAA59903.1; -;
 CC EMBL, S52227; AAB24821.1; -;
 CC EMBL, S52152; AAB24821.1; JOINED.
 CC EMBL, S52153; AAB24821.1; JOINED.
 CC EMBL, S52154; AAB24821.1; JOINED.
 CC EMBL, S52155; AAB24821.1; JOINED.
 CC EMBL, S52157; AAB24821.1; JOINED.
 CC EMBL, S52159; AAB24821.1; JOINED.
 CC EMBL, S52161; AAB24821.1; JOINED.
 CC EMBL, S52164; AAB24821.1; JOINED.
 CC EMBL, S52165; AAB24821.1; JOINED.
 CC EMBL, S52167; AAB24821.1; JOINED.
 CC EMBL, S52169; AAB24821.1; JOINED.
 CC EMBL, S52170; AAB24821.1; JOINED.
 CC EMBL, S52173; AAB24821.1; JOINED.
 CC EMBL, S52174; AAB24821.1; JOINED.
 CC EMBL, S52180; AAB24821.1; JOINED.
 CC EMBL, S52181; AAB24821.1; JOINED.
 CC EMBL, S52184; AAB24821.1; JOINED.
 CC EMBL, S52189; AAB24821.1; JOINED.
 CC EMBL, S52191; AAB24821.1; JOINED.
 CC EMBL, S52192; AAB24821.1; JOINED.
 CC EMBL, S52203; AAB24821.1; JOINED.
 CC EMBL, S52212; AAB24821.1; JOINED.
 CC EMBL, S52213; AAB24821.1; JOINED.
 CC EMBL, S52216; AAB24821.1; JOINED.
 CC EMBL, S52219; AAB24821.1; JOINED.
 CC EMBL, S52220; AAB24821.1; JOINED.
 CC EMBL, S52221; AAB24821.1; JOINED.
 CC EMBL, S52222; AAB24821.1; JOINED.
 CC EMBL, S52226; AAB24821.1; JOINED.
 CC EMBL, M76724; AAA58410.1; -;
 CC EMBL, M84477; AAA51960.1; -;
 CC EMBL, A31108; RWHL01B.
 CC PIR, A26091; A26091.
 CC PDB, 1A8X; 17-JUN-98.
 CC PDB, 1BHO; 18-NOV-98.
 CC PDB, 1BHO; 18-NOV-98.
 CC PDB, 1IDN; 25-NOV-98.
 CC PDB, 1IDO; 01-AUG-96.
 CC PDB, 1JLM; 11-JAN-97.
 CC Genew; HGNC:6149; ITGAM.
 CC MIM; 120980; -;
 CC Interpro; IPR000413; Integrin_alpha.
 CC Interpro; IPR002035; vnf_A.
 CC Pfam; PF00092; vna; 1.
 CC Pfam; PF00357; Integrin_A; 1.

CC - FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT
 CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL
 CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY
 CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.
 CC - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X
 CC ASSOCIATES WITH BETA-2.
 CC - SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC - TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
 CC GRANULOCYTES.
 CC - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC - SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC - DATABASE: NAME-PROW; NOTE-CD guide CD11c entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81695; AAA59180.1; -
 CC EMBL; Y00093; CAA68283.1; -
 CC EMBL; M29165; -; NOT_ANNOTATED_CDS.
 CC EMBL; M29487; AAA51620.1; ALT_SEQ.
 CC EMBL; M29482; AAA51620.1; JOINED.
 CC EMBL; M29483; AAA51620.1; JOINED.
 CC EMBL; M29484; AAA51620.1; JOINED.
 CC EMBL; M29485; AAA51620.1; JOINED.
 CC EMBL; M29486; AAA51620.1; JOINED.
 CC PIR; A35584; RWHUIC.
 CC HSPR; P11215; IABX.
 CC GeneW; HGNC:6152; ITGAX.
 CC MIM; 151510; -
 CC InterPro: IPR000413; Integrin_alpha.
 CC InterPro: IPR002035; VWF_A.
 CC Pfam; PF00092; vwa; 1.
 CC Pfam; PF00357; Integrin_A; 1.
 CC Pfam; PF01839; FG-GAP; 5.
 CC PRINTS; PRO1185; INTEGRINA.
 CC PRINTS; PRO0453; VWFADOMAIN.
 CC SMART; SM00191; Int_alpha; 4.
 CC SMART; SM00327; vwa; 1.
 CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 CC PROSITE; PS50234; VWFA; 1.
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 CC Signal; Magnesium; Calcium; Repeat.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 1163 INTEGRIN ALPHA-X.
 CC FT DOMAIN 20 1107 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 1108 1128 POTENTIAL.
 CC FT DOMAIN 1129 1163 CYTOPLASMIC (POTENTIAL).
 CC FT REPEAT 34 87 FG-GAP 1.
 CC FT REPEAT ? 2 FG-GAP 2.
 CC FT DOMAIN 165 351 VWFA.
 CC FT REPEAT ? 3 FG-GAP 3.
 CC FT REPEAT 402 453 FG-GAP 4.
 CC FT REPEAT 455 517 FG-GAP 5.
 CC FT REPEAT 518 576 FG-GAP 6.
 CC FT REPEAT 581 633 FG-GAP 7.
 CC FT CA_BIND 466 474 POTENTIAL.
 CC FT CA_BIND 530 538 POTENTIAL.
 CC FT CA_BIND 593 601 POTENTIAL.
 CC FT SITE 1131 1135 GEFKR MOTIF.
 CC FT DISULFID 69 76 BY SIMILARITY.
 CC FT DISULFID 108 126 BY SIMILARITY.
 CC FT DISULFID 655 712 BY SIMILARITY.
 CC FT DISULFID 771 777 BY SIMILARITY.
 CC FT DISULFID 848 863 BY SIMILARITY.

FT DISULFID 998 1022 BY SIMILARITY.
 FT DISULFID 1027 1032 BY SIMILARITY.
 FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 392 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 697 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 939 939 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 480 490 G -> A (IN REF. 2).
 FT CONFLICT 756 756 L -> D (IN REF. 2).
 SQ SEQUENCE 1163 AA; 127885 MW; 6C4E19CC3F62A473 CRC64;
 Query Match 17.2%; Score 1051; DB 1; Length 1163;
 Best Local Similarity 28.5%; Pred. No. 1.5e-67;
 Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;
 QY 13 LVFLTGLCSPP--FNLDHHPRLFPQPPAEFGYSVLQHVGGQGMVLGAPWDGSGDR 70
 DB 8 LLLFLALATSLGPNLDTBELTA--RVDSAGFGDSVVQYANS--WVVGAPQKITAAMQT 63
 QY 71 GDVYICPVGGAHNPACAGHIGDYQGNSSHP--AVNMHLGSLLETDDGGFMACAPLWS 129
 DB 64 GGLYOC---GYSTGACE-----PIGLQVPEAVNMSLIGSLASTSPSOLLACGPVH 113
 QY 130 RACGSSVFSGICARVADSPQGSAPAPRCPTV--MDVIVYLDGNSIYV--WSEYOT 186
 DB 114 HECGRNMTLTGLCFLLGPT--QITQRLPVSRQPCQEDDIFLLDSSSISRNPATMMN 172
 QY 187 FLRLVGLFLTIDPEQIQLVQYGESPVHMSLGFRTKEEVARAKNLSRREGRETKTA 246
 DB 173 FVRAVISQ--FQRP--STQSLMQFSNKQTHFFEEFRFTSNPLSLASVHQIQG--FTYTA 229
 QY 247 QAIMVACTEGSQSGRPEARLLVYTTDDESH--DGEELPALAKACAGVTRGIAVL 305
 DB 230 TAIQNVVHRLPHASGARRDATKILIVITDKKEDSDLYKDVIPMAAAGIRAIIGV 289
 QY 306 GHYLRQRPSPFLEIRITIASDPDERFFPVNTDEALTDIVDALGDRIFGLESHANE 365
 DB 290 LAFQNR-----NSKELNDIASKBQEHIFVEYEDDALQIONQKEIFALEGETTSS 344
 QY 366 SSFGLMEOIGFSTRLKXGILFGNVGAYDWGCVLMLEGGHRLPPPMALDEFFPALQ 425
 DB 345 SSFELMEOEGFSVAFTDGPVLAVGSEFTW-----SGGAFILYPPNMS-----PFI 391
 QY 426 NHA-----AYLGYSVSMILKGRRLFLSGAPFRHRRKGVIAF--QLKXDGAVRAQS 476
 DB 392 NMSQENVMDRDSYLGSTELALMKGVOSLVL--GARYOHTKAVIFTVSROW--RMKAE 448
 QY 477 LQGEQIGSYFQSELCPLPTDRDGTDLVLAAPMFLGQNKETGRVYV----- 524
 DB 449 VTGIDIGSYFQASLCSVVDTDGSTDVLIGAPHYV--EQTRGGOVSCPLPRGRMWC 506
 QY 525 --YLVGQOSLLTLOGLPEPPODARFGFAMGALPDLNQDGFADVAVGAPLEDGHQALY 582
 DB 507 DAVLVGEGG-----HPW-----GRFGALTVLGDVNGKLTVDVIGAPGEERGAIV 554
 QY 583 LYHGTS--GVRRPHAPQRIAAASMPHALSYFGRSVDGRLLDDGDLVDVAVAGQAALLLS 641
 DB 555 LFHGLGPSISPSHQRIAGSOLSRLOYFOALSGGDDLTQDGLVDLAVAGRGVLLLR 614
 QY 642 SRPIVHLPSLEVPQALISVQORDRR--GQAVCLTALCFQVTSRTP---GRWDHOFY 697
 DB 615 TRPVLMVGSWQFPAETIPSAFEBREBOVSEQTLVQSNICLYIDKRKNLIGSDLOSS 674
 QY 698 MKFTASLDWTAGAAAFDGGQRLSPRLRLSVGNVTCEQLHEFVLDTS---DYLRPVA 754
 DB 675 VTLDLALDPRGLSPRATQETKNR--SLSRVAVLGLKAKHC--NFNLLLPSCVEDSVTBIT 731
 QY 755 LTVTVALNTTKP-----GPVLESGPSTIQKLVFPSKCGCPNNECVTLVLQVNDI 807
 DB 732 LRNLFTL--VQKPLAFNLRPMLAALQRYFTASLPKXKCGADHICQDNIGISFSFP- 788

```

OY 808 RGRRAKPFVVRGGRKRVVSTLE-----NRKENAYNTSLSTIFS----- 847
DB 789 -----GLKSLVGSNLELNAEVMWVNDGDSYGT--TTFPSHAGLSRYVAE 834
OY 848 -----RULHLASLTPORESPIKVECAAPSAHARLCSGVHPVOTGAVNTLLEFEFSC 900
DB 835 GOKGQGLRSLHLTC-----DSAPV---GSGGTWSTSCRINHLIFRGGAQITFLATFVSP 886
OY 901 SSLLSQVFGKLTASSPDLSERNGLQENTNO-----TSAIOVEPHLTS----- 944
DB 887 KAVLGDRL-LLTANVSSENNTPRTSKTQLELPVKYAVYVVSSEHPTKLYLNSSESE 945
OY 945 --SESTLHRYEYHPYG--TLVPGCPGEFTTLRVQNLGCVVSGILISALL--PAVANG 998
DB 946 KESHVAMHRYQVNNIGQRLPV-----SINFWVVEVLNQEAVMWVDSVHQ 992
OY 999 NYFLSLSQVITNNASCTIVONLTPPGPPVHPELOHTNF--LNGSNTQCQVVRCHLGOL 1055
DB 993 NPSLRCS-----SEKIAPPA-SDFLAHIOKNPVLDCSIACCLFRCDVPSP 1037
OY 1056 AKGTEVSVGLRLVHNEFFRAKFKSLTVVSTFELGTEGSGVLQLTEAGRMSESLLEVQ 1115
DB 1038 SVQEBIDFLKGNLSGWRQLQKXVSVVAETTFDSVYSQLPGEAFKRA-----Q 1092
OY 1116 TRPILISLWI-----LIGSVLGLLILLALVFCIMKLGFF--AHKKIPPE 1158
DB 1093 TTVLEKYKVNHPPLVIGSSIGULLLILLITAVLYKGVFFRQYKEMWEE 1143

RESULT 13
ITAM MOUSE STANDARD; PRT; 1153 AA.
ID ITAM MOUSE P05555;
AC 01-NOV-1988 (rel. 09, Created)
DT 01-FEB-1991 (rel. 17, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Integrin alpha-M precursor (Cell surface glycoprotein MAC-1 alpha subunit) (CR-3 alpha chain) (CD11b) (leukocyte receptor MOI).
GN ITGAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88312584; PubMed=3044779;
RA Pyteia R.;
RT "Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the integrin family and an additional domain related to von Willebrand factor.";
RT Willebrand factor.";
RL EMBO J. 7:1371-1378(1988).
RN [2]
RP SEQUENCE OF 11-45 FROM N.A.
RX STRAIN=BALE/C; TISSUE=Spleen;
RX MEDLINE=86287312; PubMed=2942940;
RA Sastre L., Roman J.M., Teplow D.B., Dreyer W.J., Gee C.E.,
RA Larson R.S., Roberts T.M., Springer T.A.;
RT "A partial genomic DNA clone for the alpha subunit of the mouse complement receptor type 3 and cellular adhesion molecule Mac-1.";
RT Proc. Natl. Acad. Sci. U.S.A. 83:5644-5648 (1986).
RN [3]
RP SEQUENCE OF 17-28.
RX MEDLINE=85188276; PubMed=3887182;
RA Springer T.A., Teplow D.B., Dreyer W.J.;
RT "Sequence homology of the LFA-1 and Mac-1 leukocyte adhesion glycoproteins and unexpected relation to leukocyte interferon.";
RL Nature 314:540-542 (1985).
CC -1- FUNCTION: INTEGRIN ALPHA-M/BETA-2 IS IMPLICATED IN VARIOUS ADHESIVE INTERACTIONS OF MONOCYTES, MACROPHAGES AND GRANULOCYTES AS WELL AS IN MEDIATING THE UPTAKE OF COMPLEMENT-COATED PARTICLES. IT IS IDENTICAL WITH CR-3, THE RECEPTOR FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT. IT PROBABLY RECOGNIZES THE R-G-D

```

```

CC PEPTIDE IN C3B. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR
CC FIRINGEN, FACTOR X AND ICAM1. IT RECOGNIZES P1 AND P2 PEPTIDES
CC OF FIRINGEN GAMMA CHAIN. ALPHA-M/BETA-2 PLAY A CRITICAL ROLE IN
CC MAST CELL DEVELOPMENT AND IN IMMUNE COMPLEX-MEDIATED
CC GLOMERULONEPHRITIS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-M
CC SUBUNIT GENE DEMONSTRATE INCREASE IN NEUTROPHIL ACCUMULATION, IN
CC RESPONSE TO A IMPAIRED DEGRANULATION AND PHAGOCYTOSIS, EVENTS THAT
CC APPARENTLY ACCELERATE APOPTOSIS IN NEUTROPHILS. THESE MICE DEVELOP
CC OBESITY.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN. ALPHA-M
CC ASSOCIATES WITH BETA-2.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
CC GRANULOCYTES.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X07640; CA30479.1; -.
CC EMBL, M14293; AAA39484.1; -.
CC PIR, S00551; S00551.
CC HSSP, P1215; IABX.
CC MGD; MGI:96607; Itgam.
CC Interpro: IPR000413; Integrin_alpha.
CC Interpro: IPR002035; VWF_A.
CC Pfam; PF00092; vwa; 1.
CC Pfam; PF00357; Integrin_A; 1.
CC Pfam; PF01839; FG-GAP; 5.
CC PRINTS; PR01185; INTEGRINA.
CC PRINTS; PR00453; VWFADOMAIN.
CC SMART; SM00191; Int_alpha; 4.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; 1.
CC PROSITE; PS50234; VWFA; 1.
CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC Signal; Calcium; Repeat.
CC SIGNAL 1
CC CHAIN 17 1153
CC DOMAIN 17 1105
CC TRANSMEM 1106 1129
CC DOMAIN 1130 1153
CC REPEAT 31 84
CC REPEAT ? ?
CC DOMAIN 164 350
CC REPEAT 337 400
CC REPEAT 401 452
CC REPEAT 454 515
CC REPEAT 517 575
CC REPEAT 580 632
CC CA_BIND 465 473
CC CA_BIND 529 537
CC CA_BIND 592 600
CC SITE 1132 1136
CC DISULFID 66 73
CC DISULFID 105 123
CC DISULFID 654 711
CC DISULFID 770 776
CC DISULFID 999 1023
CC DISULFID 1028 1033
CC CARBOHYD 58 58
CC CARBOHYD 86 86
CC CARBOHYD 391 391
CC CARBOHYD 696 696
CC INTEGRIN ALPHA-M.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC FG-GAP 1.
CC FG-GAP 2.
CC VWFA.
CC FG-GAP 3.
CC FG-GAP 4.
CC FG-GAP 5.
CC FG-GAP 6.
CC FG-GAP 7.
CC POTENTIAL.
CC POTENTIAL.
CC POTENTIAL.
CC GEFPR MOTIF.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).
CC N-LINKED (GLCNAc. . .) (POTENTIAL).

```

Oy	809	GSRAAPVAVGGARRKVLVSTTLNKRKNANVTLSLIFSNLHL-----ASLTQRESPIK	864
Db	788	G-----LDLTLVVGGPQDFNMVSLVLRNDDGSDSYGTQVTVVYPSGLSYRKDSASQNPITKKPWF	844
Oy	865	VECAAPSA-----HARLCS-----VGHVPVQCAKATFLLEFEF--SCSSLSQVFGK--LTAS	914
Db	845	VKPAESSSSSEGHGALKSTTMNINHPFNPANSEVTFVTVTDVDSHAFGKLLKALVAS	904
Oy	915	SDSLERNGLTQENTQTSASVAYQYEPHLLFS--ESTL-----HREYVHPY	957
Db	905	ENMSST--H-KTKFQLELPVKYKAIYVITSDESSIRYLNFTASEMSTKVIOHQCFNNL	961
Oy	958	G--TLPPGPPEPFETTLRVQNLGCVVVSGLIISALBPVAVHAGNYFLTSLQVITNNASCI	1015
Db	962	GQRLAPV--SVFEPVPIQVNNVTWDPHPOVIFQONSSACH-----	1000
Oy	1016	VQNTPEPPEPVHP-----EELQHNRLNGSNTQCVVARCHQOLAKGTVEVGLRLVHNE	1072
Db	1001	-----TEQKSP--HSNFPDQLERTPVNCSYAVCKRIQCLDPSNTQGIIFNVTLKGNLSFD	1055
Oy	1073	FFRRAKRSLTVVSTFELGTEGGSVLQLTASRMSLSLEV--VQTRDILSLMILIGSVL	1131
Db	1056	WYITSHGHLVLVSTELIFNDPSAFALLPQESYVASKETKEVPEVHPVPLVGS	1115
Oy	1132	GGLLLLALLVPLCMKLGFPFAHKKIPEEKKEEK	1164
Db	1116	GGVLTLALLITAGLYKLGFF-----KKQYK	1139

RESULT 14

ITAE_MOUSE

ITAE_MOUSE

STANDARD;

PRT;

1167 AA.

Q06077;

01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)

15-JUN-2002 (Rel. 41, last annotation update)

Integrin alpha-E precursor (Integrin alpha M290).

ITGAE.

Mus musculus (Mouse).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.

NCBI_TaxId=10090;

[1]

SEQUENCE FROM N.A.

STRAIN=AKR;

MEDLINE=95187992; Pubmed=78821170;

Smith T.J., Ducharme L.A., Shaw S.K., Parker C.M., Brenner M.B.,

Ra Klishaw P.J., Weis J.H.;

"Murine M290 integrin expression modulated by mast cell activation";

Immunol 1:393-403(1994).

-I- FUNCTION: INTERIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT

MEDIATES ADHESION OF INTRA-EPITHELIAL T-LIMPHOCYTES TO EPITHELIAL

CELL MONOLAYERS. MICE EXPRESSING A NULL MUTATION OF THE ALPHA-E

SUBUNIT GENE EXHIBIT A MARKED REDUCTION IN THE NUMBERS OF

INTRAPITHELIAL LYMPHOCYTES IN THE GUT AND IN THE DEVELOPMENT OF

GUT-ASSOCIATED LYMPHOID AGGREGATES, SUPPORTING A SPECIFIC ROLE FOR

THIS INTEGRIN IN MEDIATING RETENTION OF LYMPHOCYTES IN THE

INTESTINAL WALL.

-I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA

SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY A

DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.

-I- SUBCELLULAR LOCATION: TYPE I membrane protein.

-I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VMFA DOMAIN. INTEGRINS

WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.

-I- SIMILARITY: BELONGS TO THE INTERIN ALPHA CHAIN FAMILY.

-I- SIMILARITY: CONTAINS 1 VMFA DOMAIN.

-I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

DR	EMBL; U12236; AAC52142.1; -.			
DR	HSDP; P11215; 1A8X.			
DR	MGD; WGI:1298377; Itgae.			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF00092; vwa; 1.			
DR	Pfam; PF00357; Integrin_A; 1.			
DR	Pfam; PF01839; FG-GAP; 4.			
DR	PRINTS; PF01185; INTEGRINA.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SMO0191; Int_alpha; 4.			
DR	SMART; SMO0327; VMA; 1.			
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.			
DR	PROSITE; PS50234; VMPA; 1.			
KW	Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;			
KW	Signal; Repeat; Magnesium;			
KW	Calcium.			
FT	SIGNAL	1	19	BY SIMILARITY.
FT	CHAIN	20	1167	INTEGRIN ALPHA-E.
FT	CHAIN	20	181	INTEGRIN ALPHA-E LIGHT CHAIN.
FT	CHAIN	183	1167	INTEGRIN ALPHA-E HEAVY CHAIN.
FT	DOMAIN	20	1114	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1115	1137	POTENTIAL.
FT	DOMAIN	1138	1167	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	?	?	FG-GAP 1.
FT	REPEAT	?	?	FG-GAP 2.
FT	DOMAIN	149	192	X-DOMAIN (EXTRA DOMAIN).
FT	DOMAIN	193	384	VMPA.
FT	REPEAT	?	?	FG-GAP 3.
FT	REPEAT	449	501	FG-GAP 4.
FT	REPEAT	503	564	FG-GAP 5.
FT	REPEAT	566	631	FG-GAP 6.
FT	REPEAT	634	686	FG-GAP 7.
FT	CA_BIND	514	522	POTENTIAL.
FT	CA_BIND	578	586	POTENTIAL.
FT	CA_BIND	646	654	POTENTIAL.
FT	DOMAIN	185	191	GLU-RICH (ACIDIC).
FT	SITE	1140	1144	GFPER MOTIF.
FT	DISULFID	72	83	BY SIMILARITY.
FT	DISULFID	130	164	BY SIMILARITY.
FT	DISULFID	698	754	BY SIMILARITY.
FT	DISULFID	814	820	BY SIMILARITY.
FT	DISULFID	884	898	BY SIMILARITY.
FT	DISULFID	998	1023	BY SIMILARITY.
FT	DISULFID	1031	1047	BY SIMILARITY.
FT	CARBOHYD	51	51	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	256	256	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	314	314	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	341	341	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	364	364	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	418	418	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	718	718	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	773	773	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	829	829	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	846	846	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	911	911	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	925	925	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	968	968	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1013	1013	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1055	1055	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1086	1086	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	1167	1167	N-LINKED (GLCNAC. . .)
FT	SEQUENCE	128983	128983	WM; B831C115DCCCFD CRC64;

QY 14 VFLTGLC-----SPFNLDHH--PRLFPGRPEAEFGYSVLQHVGAR 61

Db	4	LFHTLLCHASLKPQCAFNNVDVDMAMVTALQDPAV--LSSLHLDPSSNNQTCLLVLR	59
Qy	62	WDGEGDRGRGVYRCPCV-GAHNAFCAC-CHU---GDYOLGNSHPVNMHAGMSLLET	115
Db	60	--RSSNRNTAALYRAIISIDPELACQVEHI CMFKGYO-----GVTIV--	102
Qy	116	DODGEPMA CADLMGR-----AC-----GSSVFS9G	140
Db	103	GNHNGVLCIOYQARKFRSLNSLGTGACSLTLPNLDQAOAYFSDLEGLDPGAHVDSGD	162
Qy	141	ICARVDASFQOGSLAPTAORCPY-----MDVIVLVDGNSIYPMSEVO-----	185
Db	163	YCRSKGGS---TGEKKSARRRRIYEEDEEDGTIELAIVLDGSSIGD-SDFQAKNFI	218
Qy	186	-TFRLRLGKLFIDPEQIQVGLVQYGSFVHEWSLGDPR-----TKEEYVRAANKLSRR	238
Db	219	STMRNPFYEKCF---ECNFALVQYGAIVQIHFQDQSRDINASLAKYQSIYQVEV---	271
Qy	239	EGRETKIAQAIMVACTBEGFSQSHGRPRPAARLLVVTDGESH-DGEEPLAALKACEAGRY	297
Db	272	---TKTASAMQHVLDNFIIPSRGSRKKALKVMVLTGDIFGDPDLNTTVINSKMGV	327
Qy	298	TRYGAIVGHYLRQRDSSFLREIRTIASDBDEFNVNVDALTDIIVDALGRIGCL	357
Db	328	VFALGAVDRKDN-----NNTYRELKLIASDPKEAHTKYNYSALDGLSLKQRIYHM	382
Qy	358	EGSHAENESSFGLSEMSQIGFSTHRLKQD-ILFGMWGYDM-GGSYLM-LEGGRILFPPRM	414
Db	383	EGTVGD---ALQYQALQYGFSAQILDKQVLLGTGVAGNMSGALLYSTQNGRCFLNQT	439
Qy	415	ALDEFPALONHAAYLIGYSVSMILRGRRRLFLSGARFRRHGRKVIAFOLKKGDAVRVA	474
Db	440	AKED---SRTVOYSYLIGYSL-AVLHKAHGSIYVAGARHKLRCGAV--FELRKEDREEDA	492
Qy	475	--QSLQGEQISYRGSSELCPILDTBDCTTVLVLVAPMLFQRPQNKETRRVYVYLVCQO-S	531
Db	493	FVRRIEGQOMESYFGSVLCPVDIMDGTIDELVLAAPRY--HIRGEEGRVYVYQVPEEDA	550
Qy	532	LLTLQGTQIQQEPP--QDARFGAMGALPRLNDGDFADVVGAPLE-----DG-HOGALYLY	584
Db	551	SFSLAHTLSHGRLGTLNRFGRFMAAVGINDOKFLDVAIGARLBEPFGAGDASYSVITY	610
Qy	585	HGTOSGVARPDARQIAAASMPHALSYFGRSVDRGLDLDDGLVDVAVAGAQAIIILSSRP	644
Db	611	NGHSGGLVDSPOQIRASSVASGLHYFGMSVSGGLDFPGDGLADITVGSRSBAVVLRSRP	670
Qy	645	IYHLTPSLSEVTPQALISV---QRDCRRRGQZAVCLITAAICQOYVTSRT---PGRMDHQPY	697
Db	671	VVDLTIVSTMTFPDRLPMVFIKMDVN-----LCEEVSSVVAASEBGL--REMF	716
Qy	698	MRFTASLDEMTAGARAAPFDGSGQRSLSPRLRLSYQNVNVCQDQHLHFVLDT-----SDYLRP	752
Db	717	LNFTYDVUDVYTORQORLOCEBSSGSCSLCR-KMNGSFLCE--HFWLITBELCEBDCCSN	773
Qy	753	VALTVFALDNT---TKRGPVLNE-GSPTSIQKLVPSKDCGPNCEVTDLVLOVNDI	807
Db	774	ITIKTYTEFOYSGGRDYPNPPTLDHYKPSAIFQI--PEXQOKNKVFCIAELQILTNISQ	832
Qy	808	RGSRKARPVYRGRRKVLVSTTLENRKENAYNTSLISITFSRNHLASLTPORESPIK--V	865
Db	833	QE-----LVGVVKEVYMINISLTNCGEDSMTMTMALVYPRNLQPKKI---QKQVSDV	882
Qy	866	ECAPSAHARV-----CSVGHVPFO---TGAKVTFLEEFSCSSLSQVFGKLTASDSSL	918
Db	883	QCDPKPRASVLYMNCCKIGHPIKRGSSVNVSTIOLE-----ESVFPRKTIADITVT	933
Qy	919	ERNQTLQNTAQTSAYIQIEPHLLFSSEBTLARVEYHFGYTLPYVGGEFKITTLRVQNIQ	978
Db	934	ISNSN-EXSLARETRSLQFRHAFI---AVLSRPSV--MYMNTQSPSDHKKEFFENV---	983
Qy	979	CYVVSGLIISALLPAVAGNYFLSLQVITNNASCI-----YQNLTEPRGPV	1022


```

FT VARIANT 1041 1041 C -> S.
FT MUTAGEN 208 208 /FTID=VAR 008885.
FT MUTAGEN 316 316 D->A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 477 477 F->A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 482 482 V -> I (IN REF. 3).
FT CONFLICT 950 950 Q -> R (IN REF. 3).
FT CONFLICT 950 950 R -> W (IN REF. 3).
FT CONFLICT 1019 1019 A -> V (IN REF. 3).
SQ SEQUENCE 1179 AA; 130088 MW; E58902BDF995E1 CRC64;

Query Match 15.3%; Score 932.5; DB 1; Length 1179;
Best Local Similarity 26.8%; Pred. No. 5.4e-59;
Matches 348; Conservative 188; Mismatches 451; Indels 303; Gaps 58;

QY 8 HFLPLVPLTGLCSFNLDEHHRPLP--GPPEAEFGYSVLQH--VGGGQRWMLVGAAMD 63
DB HTLLCLASL-ALLAANVDVARPWLTPKGAAP---FVLSLHQDSTNTQTMILVTPSP--- 58
QY 64 GSGGRRGQVYRC-----PVGGAHNAFCAGKHLGDYQLGNSHPAVNMHGLMSILL 113
DB 59 -RTKRTGPGPLHRCSTLVODEILCHPV---BHVPIPKR-----HRGVTVV 98
QY 114 ETDGSGFMAACPLMSRAGS--SVFSSGICARVDAFQPGS-----LAPTA----- 159
DB 99 RS--HHGVLTICQVLVRPHSLSELTGCSLGPDLRPOAQANFPDLNLDPDPAVDVT 156
QY 160 -----ORCPYMDVIVLIDGNSIYP- 180
DB 157 GDCYSKEGGEDDVNTARQRALKEKEBEDEBEDEBEAGTEIAIILDGSSGIDPP 216
QY 181 -MSEVOTFLRLVGLKFLIDPEQIOVGLVOYGESPVHEWGLGDFRTKEEVVRAKILSRRE 239
DB 217 DFORADDFISNMNRNFYKCFECNFAIVQYGVITQTEFLRDSQDMAGLAVQNT-QY 275
QY 240 GRETKAOLIVACTGEGFSQSHGGRPEARLLVYVTDGE--SHDGEELPAALKACEAGRYT 298
DB 276 GSVTKTASMOHVLDISITSSHGRKAKKVMVLDGGIFEDPLMTTVINSPKMQGVE 335
QY 299 RYGIATLGHYLRORDPSSFLREIRTIASDPDERFFENVTDALTDIVDALGDRIFGLE 358
DB 336 REAIGV-GEFYSART---ARELNIAADPBETHAKTYNMALDGLSKLRYNISME 390
QY 359 GSHAENESSFGLEMSOIGFSTRKLD-GILFGMVGAYDMGGSVLMLEGHRLPPPRMA-- 415
DB 391 GTVGD---ALHYQLAQIGFSAQIILDERQVLLGAVGAFDWSGAL-----LYDTRSRRG 440
QY 416 --LEDEFPALQNHAA---YLGYSVSSMLLRGRRRLFLSGARFRHRKVIAPOLKKG- 469
DB 441 RFLNQTAAADAADAEEAAQYSLGYAV-AVLHKTCSLSYVAGAPQYKHGAV--FELQKEGR 497
QY 470 AVRVAQSLQGEQIGSYFSGSELCPDLTDRTDVTDLVLAAPMFLGPQNKETGRVYVYLVQ 529
DB 498 EASFLPLVLEGBQMGSYFSGSELCPVDIDMGSTDFLLVAAPFY--HYHGEGRVYVYRLE 555
QY 530 Q-SLLTLOGLQPEPP-QDARFGFANGALPDINODGFADVAVGAPLE-----DGHQ-QAL 581
DB 556 QDGSFSLARILSGHCFGTNARFGFAMAAGDLSQDKLTVAICAPLEGFAGADDGASFGSV 615
QY 582 YLYHGTQSGVRPHPRQRIAAASMPHALLSYFGRSVDCRLDLDGDDVLVDVAVGAQAAILLS 641
DB 616 YLYNGHMDLSPSPQRIRASTAVAPGLQYFGMSMAGGFDISGDLADITVGTGQAVFR 675
QY 642 SRPIVLTPLSLVTPQALSVVORDCRRRQEAVALCTLAALCFQVTSRTPGAMD--HOFVVR 699
DB 676 SRPVVRLKYSMAFTSALPI-----GFNGV-VNVRLCFEISSVTTTBSGLAREALLN 726
QY 700 FTASLD-----EYTAGARAAFDGSGQRLSPRRRLSVGNVTCQ 738
DB 727 FTLDVDVGQRRLQCDVRSCLGCLREWS-----SGSQLCEDLLIMPTEGELCE 777
QY 739 LHFHVLDTSDYLRPVALTVTPALD---NTTKGPVLNE-GSPTSIQKLVPSKDCGPRDN 793
DB 778 -----DCEFSNASVKSIVQLQTPREGQTDHPQTLDRYTEFPALFQI-PYEKACKNKL 827

```

```

QY 794 ECVTDVLQVNMIDIRSKRKAPEVVGRRKVLVSTLLENRKENAYNTSLIIFSRNLHA 853
DB 828 FCVAELQIATVTSQGE-----LVVGLTKELTMINILNNSGDSYMTSMALNYPNLTQK 881
QY 854 SLTPRESPIKVECAAPSAHARL---CSVGHDPV-QGAKVTFLLREFPGSSLLSOVF 908
DB 882 RM--OKPPSPNQCDPPQPVASVLIMNCRIGHPLVKRSSAHVSVMQLE-----ENAF 932
QY 909 GULTASDLSLRNGTLOENTAQTSAYIOY-----EPHLF--SSESTLHRYE--VH 955
DB 933 PKRTDITVYVYVTSNERSLANETHTLQFRHGFVAVLSPSIMVYVNTQGLSHKKEFLPH 992
QY 956 PVGTLVPGRPEFTKTLRVQNLGCVVSGLIISALLPAVAGNYFLSLQVIT----- 1009
DB 993 VHGENLRF--GAEYQIQCVP---KLNGLOVAANKLRTQASIVCTWGSQERACAYSSV 1046
QY 1010 -----NNASCTV---ONLTPEPQPVHPE-ELQHTNRLNGSNTQCCVVRCHQLAKG 1058
DB 1047 QHVEEMHSVSCVIASDKENVT-----VAAEISMDHSEELLKDYVTELOI---LGEISFN 1096
QY 1059 TEVSQGLRLVHNE-----FPRAKFKSLTVVSTFELGTEGGSVLQLEASRWSSESLLEV 1113
DB 1097 KSLYEGLAENHRRTKITVFLDEKYNHL----- 1125
QY 1114 VQTRPILISLWILIGSVLGLLLALLVFCIMKLGFFAHL 1153
DB 1126 -----PIIK-----GSV-GGLLVILVILVIFKCGFFARK 1155.

```

Search completed: July 16, 2003, 07:52:48
Job time : 36.365 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 / Search time 59.9093 Seconds
(without alignments)
1872.646 Million cell updates/sec

Title: US-09-647-544-2
Perfect score: 6106
Sequence: 1 MELPFTVTLPLVFLTGLC.....GFFAHKKIPREEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	ID	Description
1	1894.5	31.0	1180	2 A35854	integrin alpha-1 c
2	1887	30.9	1151	2 A45226	integrin alpha-1 c
3	1756	28.8	1181	2 A33998	integrin alpha-2 c
4	1754	28.7	1170	2 A45914	integrin alpha 2 s
5	1744	28.6	1178	2 S44142	VLA-2 protein homo
6	1105	18.1	1170	2 S03308	cell surface glyco
7	1097	18.0	1163	2 I56126	lymphocyte function
8	1085.5	17.8	1153	1 RWHUB	cell surface glyco
9	1054	17.3	1163	1 RWHUB	cell surface glyco
10	1022.5	16.7	1153	2 S00551	leukocyte surface
11	932.5	15.3	1179	2 A53213	integrin alpha-E c
12	738	12.1	1035	2 I58409	integrin alpha-9 c
13	732.5	12.0	1054	2 JG7294	alphan integrin -
14	690	11.3	1039	2 A41131	lymphocyte-Feyers
15	684	11.2	1038	2 S06046	integrin alpha-4 c
16	661	10.8	1041	2 T31437	integrin alpha-1 c
17	655	10.7	1041	2 A55348	integrin alpha-1 c
18	640.5	10.5	1137	2 JCS950	integrin alpha-3 c
19	602.5	9.9	1051	2 A40021	integrin alpha-6 c
20	600	9.8	1091	2 A41543	integrin alpha-6 c
21	598.5	9.8	1072	2 B36429	integrin alpha-6 c
22	596.5	9.8	1072	2 A38457	integrin alpha-6 c
23	594.5	9.7	1135	2 T16186	alpha-7 integrin -
24	584.5	9.6	1044	2 T10050	integrin alpha-v c
25	581.5	9.5	1053	2 I55514	VLA-3 alpha subuni
26	576.5	9.4	1044	2 S16516	integrin alpha-8 c
27	576.5	9.4	1051	2 A35761	cell surface glyco
28	574	9.4	1048	2 A27421	integrin alpha-5 c
29	573.5	9.4	1106	2 S38783	integrin alpha cha

30	570	9.3	1045	2 S60571	integrin alpha v c
31	566	9.3	1053	2 S44250	integrin alpha-5 c
32	563.5	9.2	1034	2 A36108	integrin alpha-V c
33	562	9.2	1049	2 A27079	fibronectin recept
34	543	8.9	1146	2 S40311	integrin - fruit f
35	540	8.8	1039	2 A34269	integrin alpha-2b
36	495	8.1	1037	2 A60163	glycoprotein IIb -
37	485	7.9	1226	2 S44824	F54F2.1 protein -
38	475	7.8	1394	2 A29677	position-specific
39	445.5	7.3	1139	2 S28277	hypothetical prote
40	438	7.2	1086	2 T18523	integrin alpha cha
41	425.5	7.0	1115	2 T09403	integrin alpha cha
42	425.5	7.0	1115	2 T09433	integrin alpha cha
43	394.5	6.5	191	2 I47230	VLA-2 protein - pi
44	374	6.1	764	2 I36916	glycoprotein IIb -
45	303.5	5.0	604	2 I36917	glycoprotein IIb -

ALIGNMENTS

RESULT 1					
A35854					
Integrin alpha-1 chain precursor - rat					
C:Species: Rattus norvegicus (Norway rat)					
C>Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 20-Sep-1999					
C:Accession: A35854; S11243					
R:Rignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, U. Cell Biol. 111, 709-720, 1990					
A:Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin a					
A:Reference number: A35854; MUID:90338125; PMID:2380249					
A:Accession: A35854					
A:Status: preliminary					
A:Molecule type: mRNA					
A:Residues: 1-1180 <IGN>					
A:Cross-references: GB:X52140; NID:G56493; PIDN:CAA36394.1; PID:G56494					
A:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology					
A:Keywords: cell adhesion; cytoskeleton; transmembrane protein					
F:170-345/Domain: von Willebrand factor type A repeat homology <vWA2>					
Query Match					
Best Local Similarity 31.0%; Score 1894.5; DB 2; Length 1180;					
Best Local Similarity 36.5%; Pred. No. 2.5e-133;					
Matches 436; Conservative 221; Mismatches 461; Indels 77; Gaps 24;					
QY	13	LVPLTGLCPFLNDBEHRLFPPEPEAFEGSVLOHVGCGGQRMVLVGA PMDGPSSDRRGD	72		
DB	19	LTVLIGFVCSFNVDVKMSMSFGVEDMFGITVQYENEBKRWLIGSPLVGPKARTGD	78		
QY	73	VYRCPEVGAHNAAPCAKHLGDYQLGNSHPAV-----NMHLGMSLLETGDPGFWACAPL	127		
DB	79	VYKCPVGERAMPCKVDLP-----VNTSIPNVTEIKENMTFG-STIVATNPNGFLACGPL	133		
QY	128	WSRACGSSVSSGICARVDASFQPGSLAPAGRCPTMYDVIVYLDGNSITYPMSSEVOTF	187		
DB	134	YAYVCGHLHYTTGICSDVSPFTFYVNSFEAP-VQECSTQLDIVYLDGNSITYPMSVIAF	192		
QY	188	LRLVIGKLFIDPELOVGLVYGSPVHMSLGPFRTEEVVRAKXNSRESEGETKTAQ	247		
DB	193	LNDLKRDDIDPKQYQVIGVQYGENVTHEFPLNKSYSTEEVLVANKIGRGGLQTMFAL	252		
QY	248	AIMVACTEGFSQSHGGRPEEARLLVVYTDGSHDGEELPAALKACEAGRVRYGIAYLVGH	307		
DB	253	GIDPARKEAFTFARGARGVAKVWVITYDGSHPNRYKQVIOCEEDNIRFSALIGH	312		
QY	308	YLRQRDPSSFLREIRTIASDPDERFFPNVYTDAAALTDIVDALGDRIFGLGSHANESS	367		
DB	313	YNRGNLSTEFVEERIKSIASEPTEKHFFNVSDELALVTIVVALGERIFALPATADQSAAS	372		
QY	368	FGLMSQIGFSTHLKQGLLGWVGAYDMGGSVLMLEGGHRLFPFPMALDEEPALQNH	427		
DB	373	FEMENSOTGFSAHYSQDMVLMGAYGADWNGSTVWQKANGVNIHNTTPTQTE--PAKNE	430		
QY	428	--AAVLGYSVSMLRGGRRLFLSGAPRFRRHGVIAFQLKKGDAVVAQSLQGEQIGSY	485		

```

Db      431 PLASYLGVTNSATIPGD--VLYIAGQPRVNHGTQVVIYKKA-EDGNINIIQTLCGEQIGSY 488
Qy      486 FGSLECLPDLTRDCTTDVLLVAAPMFLGPONKKTGVVYVYLVGQOQLTLTQGLQF----- 541
Db      489 FGSVLTITDIDKOSYTDLLVGAAPMWTGKEQGVYVAV--NQTREYQMSLEPIROT 547
Qy      542 -----BPQDARFGFAMGALPDINOGFADVAAGVAPLEDHOGALYLYHG 586
Db      548 CCSGLNDNSTCKENKNEPCGARFGTAAVAKD.LNVGDFNDVIGALLEDHAGAVIYHG 607
Qy      587 TOSGVRPHPAQRIAAASMPHALSYFGSSVDGRJLDGDDI.LVDAVAGAAILLSRPV 646
Db      608 SGTITIEAAYQRI.PSGGDDGKTLKFQOSIHGEMDLNGDGLTDTYIGLGAALFMARDVA 667
Qy      647 HLPLSEVTPQAISVVOQDRCRRGQAVCLTALCQVYSTRPGRDHOFYMFATSLDE 706
Db      668 VAVATNPFBNKVNICKKCRVGEKTCINATMCHVAKLSKEDSIYEADLOYRVTLDS 727
Qy      707 WTGAAAPAFDGSQGLSPRLRLSVGNVTCQQLHFLVDTSDYLRLPVALTVTFALDNTTK 766
Db      728 LRQISLSPFGQTERKIQF--NITVASECIRHSFYMLDKHDFQDSVRTLDF--NLID 782
Qy      767 P--GVLNBSGSPSIOKLVFPGKQDNECVTDVLQVNMDIRSKKAPFVVRGGRRV 824
Db      783 PENGVPVLDLALPNSVHEHIFPAKDCNKKERCISDLTLNVST--TEKSLLIVKSQHDRF 838
Qy      825 LVSTLENKKNVNTSLSIIFSRNLHLASLTP--QRESPIKVECAPSAHARLCSGHV 883
Db      839 NVSLTKNKGDSVYNTRTVQHSFNLIFSGIEIIOQDS-----CESNONITCRGYPR 891
Qy      884 FQTGAQVTELEFEFCSSLSLQVFGKLTASSDSLERNGLQENTAQTSAYIOYEPHLF 943
Db      892 LRAGEIVTETKLIQFQNTSHSENAIHLHLSATSDSEEPLESINDENVNIGIPVKEVGLQF 951
Qy      944 SSESITLHRYEHPDYGLP-----VGPGEPEKTLTAVQNLGCVVSGLIISALLP-AVAV 996
Db      952 YSSASHSHTSVANETIPETINSTEDIGNEINVFYIRKGFHPMDELSTISFPLMTAD 1011
Qy      997 GGNVFLSLQVLTNNASCIVONLTBPG-----PVPHEPELOHTRLNGSNQCCVAVC 1050
Db      1012 GYPLVLYPGWSSSDVNCPRSLDEDFGINSKKMTISSEVUKRTIQDCSTCGVAT 1071
Qy      1051 HLGLQAKG--TEVSGVLRLVHNEFFRRAKFKSLTVSTELGTEESVQLTEASHWSBS 1109
Db      1072 TCSLPSDSLQVAVSL--LLMKPTFIRAHFSSLNLTLRGEIKSENS--LTLSSNKKRL 1128
Qy      1110 LLEAVQT-RPLISLMTLIGSVGLLALLLVFCMLKLGFAHKKIPEEKREE 1163
Db      1129 AIQISKDGLPGRVPLVWILSAFAGLLMLLMLLALMKIGFF--KRPKKKMKK 1180

RESULT 2
A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #ext_change 31-Dec-2000
C:Accession: A45226
R:Blakeschewitz, R.; Epstein, M.R.; Marcantonio, E.E.
U: Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; MUID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Experimental source: hepatoblastoma cell line HepG2
A:Note: sequence extracted from NCBI backbone (NCBI:124326)
C:Superfamily: unassigned
C:Domain: von Willebrand factor type A repeat homology <WAL>
F:142-317/Domain: von Willebrand factor type A repeat homology <WAL>

Query Match      30.9%; Score 1887; DB 2; Length 1151;
Best Local Similarity 36.3%; Pred. No. 8.8e-133;

```

```

Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
Qy      23 FNLDEHPLRLPFGPEAREGYSVLQHVGGGRWMLVGA.PMIDPSGDRGRDVRCPVGAH 82
Db      1 FNVADKSNKSTSGPVEDMFGITVQOYENEBGKMWLLGSLYGPKNRTGDTVKCVGGE 60
Qy      83 NAPCAKGLDGYOLGNSSHPAV-----NMHLGMSLLETDGDFNACAPLWSRACSSVF 137
Db      61 SLPCVKTDLR-----VNTSIPNTEVKENMTG--SLTVNPNNGFLACGFLVYRCGLHY 115
Qy      138 SSGTCARVADASFQPGSLAPTAQRCPTVMDVYIVLDGNSNIYPMSEVQTFRLRVGLFI 197
Db      116 TTGISDSVPFQVNVNSIAP--VQECSTQLDIYIVLDGNSNIYPMQSVAFNLDLKRD 174
Qy      198 DEQIYGVQVQGESPVHMSIGDRTKEEVRAKNLSREGRETXTAOAIMVACTGCF 257
Db      175 GPKQTOVGIVQYGEVNTHEPRLNKSSSTBEVLVAAKKIVQKGRQTMALGLDTARKAF 234
Qy      258 SOSHGGRPEARLLVVVTDGESHGDEELPAALKACEAGRVTRYCIAVLGHYLRQRDPS 317
Db      235 TEARGARGVKKVWVIVTDGESHDRHLKVIYQDCEDENIQFSAIGSVNRGLSTEK 294
Qy      318 FLREIRTLASDDEFFPNVYDDEALTDIVDALGPRTIGLBSHAENSSFGLEMSQGF 377
Db      295 FVEEIKSISASEPTEGHFNVSDDELAVITVTKLGRIPALATADQSAASFEMESQGF 354
Qy      378 STHRLKDGILFCMGVAYDMGSLVWL.BEGHRLFPFRLMLDEFPALONHAAYLGYSVS 437
Db      355 SAHYSQDMVMAVCAVATYMNQVWOKASQIIPNTTFNVBSTKNEPLASYLGITYNS 414
Qy      438 MLRGARLFLSGARFRHGRKVIAFOLKQDGAHVAAQLOGEOIGSYFGELECLDTR 497
Db      415 ATASSGDVLYIAGQRYVNHGTQVLIYRM--EDGNIKIQLTSGEQYFGSILTTTIDK 473
Qy      498 DQTTVLLVAAPMFLGPONKKTGVVYVYLVGQOQLTLTQGLQF----- 541
Db      474 DSNTDILLVGAAPMWTGKEQGVYVAV--NQTREYQMSLEPIKOTCCSRQNSCTT 532
Qy      542 -----BPQDARFGFAMGALPDINOGFADVAAGVAPLEDHOGALYLYHGTOGVRPHAPR 598
Db      533 EKNEPCGARFGTAAVAKD.LNVGDFNDVIGALPLEDHGAVIYHSSGKTIKREYQF 592
Qy      599 IAAASMPHALSYFGSSVDGRJLDGDDI.LVDAVAGAAILLSRPVHLTPSLSEVTPQA 658
Db      593 IPSGGDGKTLKFQOSIHGEMDLNGDGLTDTYIGLGAALFWSHDAVAVKTMFBNK 652
Qy      659 ISVVRDRCRRGQEAVALCTALCQVYSTRTEGRMDHOFYMFATSLDEMTAGARAFQGS 718
Db      653 VNIQKQCHMEGKETVCINATVCEVUKLSKEDTTEADLOYRVTLDSLRQISRSFSGT 712
Qy      719 GORLSPRLRLSVGNVTCQQLHFLVDTSDYLRLPVALTVTFALDNTTKP--GPVINEGSP 776
Db      713 QERKQFQ--NITVASECTKHSFYMLDKHDFQDSVRTLDF--NLTPENGVPVLDLSP 767
Qy      777 TSIQTLVPSKDCGPDNECVTDVLQVNMDIRSKKAPFVVRGGRRKVLVSTLENREN 836
Db      768 NSVHEYIPFAXDCNKKECISDLSLHV--ATEKDLIIVRSQDKFNVSILTVNXTDS 823
Qy      837 AYNLSLTIIFSRNLHLASLTPORESPIVVECAPSAHARLSYGHFPVOTAKVTFLEF 896
Db      824 AYNNTTIVHSPNVLVFSI-----EAIQDSC--ESNHNITKQVGPFLRQGMVTFKILF 877
Qy      897 EFGCSLSLQVFGKLTASSDSLERNGLQENTAQTSAYIOYEPHLFSSSESTLHRYEHP 956
Db      878 QPNTSYLMEVNTIYISATSDSEEPETISDVNVANISIVKKEVGIQFSSASEVHISTAA 937
Qy      957 YGTLP-----VGPGEPEKTLTAVQNLGCVVSGLIISALLPAVAGN-----YFLSLQV 1007
Db      938 NETVEVINSTEDIGNEINIFYLRKSSGFPMPPELKSISFPNMTSNQVPLVPTGLSS- 996
Qy      1008 ITNNASCIVONLTP-----PQPVPEPELOHTRLNGSNTOCCVAVCHGLQAKKTEV 1061
Db      997 -SENANCRPHIFEDPFSINSKKMTTSTDLKRGITLDCNTCKFATINCNTLS--SDISQV 1054

```

Oy 1062 VGLLRLAHNNEPRFAKRSILTVVSTFELGVEEGSVLTQTSAMSESLAEVQT-RPIL 1120
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1055 NVSL--IIMKPFFIKSYSSLSMLTTIRGELRENNNS-LWSSNQRELAIQISKDLPGR 1111

Oy 1121 ISLWTIGSVLGILLALLLVFCMKLGFPFHAKKIPEEKKEE 1163
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1112 VPLMWTLISAFAGLLLMKLITLALMKIGF---KRPLKKMEK 1151

RESULT 3
A33998
integrin alpha-2 chain precursor - human
M.Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ct
C.Species: Homo sapiens (man)
C.Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 21-Jul-2000
J.Accession: A33998; S56793; A53117
R.Takada, Y.; Hemler, M.E.
J. Cell Biol. 109, 397-407, 1989
A.Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit [platele
A.Reference number: A33998; PMID:89308879; PMID:2545729
A.Accession: A33998
A.Molecule type: mRNA
A.Residues: 1-1181 <TAR>
A.Cross-references: GB:X17033; NID:G33906; PIDN:CAA24894.1; PID:G33907
A.Note: The authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
Biochem. J. 279, 419-425, 1991
R.Carimeli, B.; Parmenier, S.; Leung, L.L.; McGregor, J.L.
A>Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIC*, GPIa an
A.Reference number: A56793; PMID:92061944; PMID:1953640
A.Accession: S56793
A.Molecule type: protein
A.Residues: 30-43 <CAR>
A.Experimental source: platelet
R.Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A>Title: The human alpha-2 integrin gene promoter. Identification of positive and negativ
A.Reference number: A53117; PMID:94103255; PMID:8276836
A.Accession: A53117
A.Molecule type: DNA
A.Residues: 1-16, 'V', 18-21 <ZUT>
A.Cross-references: GB:L24121; NID:G400342; PIDN:AAL16619.2; PID:G4583535
A.Note: authors translated the codon GTA for residue 17 as Ieu
C.Genetics:
A.Gene: GDB:ITGA2; CD49B
A.Cross-references: GDB:128031; OMIM:192974
A.Map position: SqII.1-SqII.2
C.Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
C.Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
F.I-29/Domain: signal sequence #status predicted <SIG>
F.F30-I193/Domain: extracellular #status predicted <EXT>
F.I172-347/Domain: von Willebrand factor type A repeat homology <WA2>
F.I114-I154/Domain: transmembrane #status predicted <TM>
F.I155-I181/Domain: intracellular #status predicted <CYT>
F.I105,I112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent)

Query Match 28.8%; Score 1756; DB 2; Length 1181;
Best local Similarity 34.3%; Pred.No. 6.2e-123;
Matches 418; Conservative 219; Mismatches 455; Indels 126; Gaps 31;

Oy 11 LPVLVTDL-----CSFNLDHHPRLFPGPPBEAFGEYGVLOHVGGGQRMMVGAPWD 63
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 11 LPLLVLVALSGILNCCLAVNVGRLPEARIFGFSPSEQGYVAQGPFINKGMWLVSFPM 70

Oy 64 GPSGDRCDDVRYCPYGGAHNAPCAKGHL-GDYOLGNSSHPAVNPHLGMSLLETDGCGFM 122
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 71 GFPEHRMDVVYKCPCV-DISTATCEKLNLOTSTSIPTVMETKTNMNLGIILTRNMGTCGFL 129

Oy 123 ACAPLMSASGSVSSSFGICARVDASQPQGSILAFTARCCTTYDVIVULDGSNISIPWS 182
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 130 TCGPLMAQQCKQNYTTTGCDISPDQLSASFSAITOPCCSLIDVVVVVCESNISIPWD 189

Oy 183 EVQVFRLRVLNGKLFDIDPEQIQVGLVOGESPVHEWSLDFPTKEEVVRAANKLRRRGRE 242

Db	190	AVKMFLEKVEQSLDIDFTPTQVGLQYANPRVAVFNLTYYTKKEEMI VAIISQISQYGGD	249
Qy	243	TKTAQAIMVACTEGFSQSHGRPEARLLVWVTDESHDEBELPAALKACEAGRVTYGI	3020
Db	250	TNTFGAIQYARKVAYSAASGRRSAPTKVMVVVTDESHDSMLAAVIDQCHNDILAFGI	3090
Qy	303	AVUGHYLRORDDSSFLREITLITLADDPDRFFPNVTDEALTDIYDALGRI FGLBESHA	3626
Db	310	AVLGYUNRNALDTPKNLKEIKAIASIPERYEFNVDSDEALLERAGLTGGEIPISECT -V	3686
Qy	363	ENSSSFGLEMSQIGFT--HRKDGILFGMGAYOMGWSYLM--LEGGHRLJPPPMALDE	419
Db	369	QGGDNTPQEMESQVGFADYSSQNDILIMGAVCAFEMSGTIVOKTSHGLLFP-----KOA	4233
Qy	420	FPBALQ--NHAALGYSVSMRLGGRRLFLUSGAFRPHRKGVALFOLKXDAVRVAQSL	4773
Db	424	FDQILDRHNSSTVIGSVNA--ISTBESHFVAGAPRANYTQIYLVSGNEGNITVIOAH	4824
Qy	478	QGEQIGSYFSGELCPJLDTDGTDTVULLVAAPMLGPNKETGRVYVY---LVGQSL	5322
Db	483	RGDQIGSYFSGVLCSDVDVKDTITTVULLVGAAPMWSDLKEEGRVYVFTTIKKGILGHO	5424
Qy	533	LTIQGLPDRPPDPAFAFGMALDILNQDGAIVAGAPLJEDHGQALYIYHGTOSGR	5929
Db	543	--LEG---PEGIENTFGAIALALDIMDGENDVIVGSPLENOGSAVYIYNHQGTIR	5973
Qy	593	PHAPORIAAA--SMFALSYFGRSDGRJLDJDDDLVVAAGAQAAILSSRPVILTP	6506
Db	598	TKYSQKILSDGDAFRHLOYFRSLDGTGDLNGDSITVSIAGAGQVQYOLMSQSIADVAI	6555
Qy	651	SLEVTPOAISVYORDCRRRGDEAVCLTALCEQVTSRTPGRMDHQFYMRFTFASLDEWTAG	710
Db	658	EASFTPEKITLVKNAQ-----IILKLGSAKFR-PTKONQVAIVANITLDA----	704
Qy	711	ARAPFGSGORLSPRLL-----RLSVGNV-----TEQULHFVILDTSDYLRVALVT	756
Db	705	-----DGFSSRVTSRGLFKENNERCLOKRMVNVNAQSCPEHIIYIOEPSDVVMSLDRVD	759
Qy	759	FALDNTTKGC--PVLNBSPTSIOQLVFPSPKDCGPNCEVTDLYQNMDRGSRKAPV	816
Db	760	ISLEN--PETSBALEAVSETAKVSIPIPHKDCGEGDGLCISDVLVDR-QIPAAQEOPTI	815
Qy	817	VRGRRKVLVSTLLENRKENAYNTSLIIFSRNHLASTLTPQRESPI---KVEC-AAPSA	8727
Db	816	VSNQNRILFPSTVTLKKKRESAVNTGIVDFSENLFPSFS---LPVDGTEVTCQVAAQ	8711
Qy	873	HARLCVGHVFPOTGAKVTFLLLEFFSCSSLISQVFGKLTASDLSLERNGTLQENTAO	9322
Db	872	KSAVCQVGPALRREQOVFTFINPFDNLONQNASLSFQALSSEOEENKA--DNLVNLK	929
Qy	933	AYIQYERPHLLFSEESTLHREYHPRGTLP-----VGRPEFKTTLAVONIGCVVSGLI	966
Db	930	IPLLYPAEHLTRSTININFEIISDGNVPSIVHSEFVGPFTESLKV-TTGSVPVSMAT	9688
Qy	987	ISALLPVAHAGNVYFLSIOVTNNASCIVONLTPP-----GPVHPEELQHTNRL	10303
Db	989	VIIHIFQYIKENKPLMYLTGVOTDRAGDISCNADINPLKIGQTSISSVFSKSENRRHREL	104
Qy	1039	NGSNTQCOVVRCHLQOLAKGTSEVSGLLRLVHNEFFRAKFSLTJVSTELGTEGSQL	109
Db	1049	NCRTASCNSVTCWLKDVHMKGEYFVNVTTRILWNGTFASSTQYQVOLAABAINTYNPEIY	110
Qy	1099	QLTBASRMSSELLEVQTPRILISMT-----LIGSVLGGLLALLALVYCLW	114
Db	1109	-----VIEDNTVTIPIIMPKPDEKAEVPTGVLIGSIAGILLLALVALW	115
Qy	1146	KLGFPAAK-----KIPBE 1158	
Db	1155	KLGFPKAKYKTKNPDE 1172	

145914
 integrin alpha 2 subunit - bovine (fragment)
 C.Species: Bos primigenius taurus (cattle)
 C.Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 20-Sep-1999
 C.Accession: I45914
 R.Kamata, T.; Puzon, W.; Takada, Y.
 J. Biol. Chem. 269, 9659-9663, 1994
 A.Title: Identification of putative ligand binding sites within the I-domain of integrin
 A.Reference number: A54402; PMID:94193647; PMID:7511592
 A.Accession: I45914
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1-1170 <KAM>
 A.Cross-references: GB:I25886; NID:9439695; PIDN:AAB59255.1; PID:9439696
 C.Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
 F161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.7%; Score 1754; DB 2; Length 1170;

Best Local Similarity 33.9%; Pred. No. 8.6e-123;

Matches 411; Conservative 235; Mismatches 465; Indels 102; Gaps 30;

```

9  LFLPLVFLTLGL---GSPNLDEHRLFPPEPEAFGYSVLQHVGGGGRMMVLGAPWDGP 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2  LQVLVFSOGILNCVAAYVNGLPKAKIFSGPSSEQGYAVQVQFINDKGMVLVGSFWSGP 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66  SGDRGDVYRCPCVGAHANAPCAKGH- GDYOLGNSHPAVNMHLSLLETDDGGGFMAC 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62  PKRMMDVYKCPV-DLSTTCEKLNQTSMSNVTDEKTNLSGLTLIRNVTGGFLTC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125  APLMRACGSSVFSSGICARVDASFPQGS LAPTAQRCPPTMDVIVLDSGNSIYPMSEV 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121  GPLMAQCGSQYVTTGVCSDPDQLRTSFAPAVQTCPSFIDVVVVCDSNSIYPMDAV 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
185  QTFRLRVKGLFIDPEQIOVGLVQGESPVHEWSLQDFRTKEVVAANKLSRERETK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181  KNFLKLVQGLDGPRTKTMGLIQYANPRVFNLTFFSKBEMIKATSQTFQYGGDLTN 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245  TQOAIIVACTEGFSQSHGGRPEARLLVVTTDESHDEGLPALKACAGATRYGIAY 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241  TFRALTYARDTAVSTAAGRPKATKMYVTTDESHDGLKAVIDQCKNDLIRGIAY 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305  LGHYLRORQDPFSLEIRTIASDPDERFFVNTDEAALTDIALDGRIFGLESHAEN 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301  LGVNNALADTKNLIKRIKAIASIPTRHFVNVSDEADLLEKAGTIGEQFISIEGT- 359
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365  ESSFGLMSGIOGSTRRLDGL-ILRGMCAYWGSVLM-LEGHRLPPRALDEPR 421
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
360  GDNFQWEMSQVGSABYSQNNILMLGAGCAYWGSCTVQKTPHGLIFS-----KQAF 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422  PALQ--NHAAYLGYSVSMILRGGRFLTSGAPRFRHGRVIAFOKKQCAVVAQSLQ 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
415  QILQDPNHSYLCYVAAS-ISTGNSVHFVAGAPRAYTQGIYVSNENGNVTIYOSQ 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480  EQIGSYFSGELCPDLTRDGTDTLLVAAMPFLGPONKETGRVYVYLVGQSLTLTQ 539
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
474  DQGSYFGSVLCAVDNKTITITDVLVGAAPMYNMDLKEGGRVYLFITKGIIMHGF 533
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
540  QEPEDDAFFRPMGLPDLNODGPDVAVGAPLEHGQCALYLYGTOSGVRPHQAQ 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
534  GNGLENARFGSIALSLIMNDGFNDVIVGSPLFNQNSGAVYIYNGHGMRLRYSKI 593
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
600  AAASMPHA--LSYFGSVYDGRDLDDDDVAVVAGQAAILLSRPVYHLFTSLEVT 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
594  LGSDFRFSHLYQFSGSLDGYDLNDSITDVSVGFQGVVQVWMSIADVSDAFTRK 653
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658  AIVSVQRDRRRRQGEAVCLTALCFQVTSRTPEGRMDHGFYMRFTASLDEMTAG 713
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
654  KITLKNKAE-----IKLKCFSAKFR-PTNQNNUVAIYVNIITIDQFSSRVIS 704
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
714  APDGSQRLSPRLRTSVGNVTCQHLHFVLDTSYLRPALTLVTALDNTTKRG--PV 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
705  LFKENNERCLQKTMIVSOAO-RCSEYI IHIQEPSDIISPLNLCMNISLEN--PG 760
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 772 NEGSPTSIQKLVFPSKDCGPDNECVTDLVQVNMDIRSKAPFVVRGRKVLVSTLE 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 761 EAYSTVAVFSLFPHKDGDDGVCISDLVAV-QQLPATQOQPFIVSNQNKRLTFSV 819
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 832 NRKENAVNTSLIITSRLHLASLTPORESP---KVEC-AAPSAHALCSVGHVFO 887
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 820 NKKEAVNTETIIVDSENLFFASWS---MPVDGETVCOJASQKSVTCNVGPA 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 888 AVVTLLPEPEFCSSLSQVGVKLTASSDSLEKNTGLQENAQTSAYIQYEPHLLF 947
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 876 QCVTTTFIDFVLQNLQNASISFRALSESQEN--MDNSVNLKSLIYDEHIT 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 948 TLHRYE-----VAPYGTLPVGPPEFRTTLRVONLGCYVSGIISALP 997
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 934 NINPFVSLDGVSSVHVSFEDI---GPKFISIKV-TTGSVPVSMASVHIHQY 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 998 GNYFLSLSOVITNNA--SCIVQNLTEPPGPRVHP-----EELQHNRLNG 1040
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 989 KNPLMYLTGVHTDQAGDISCEAE-----INPLKIGOTSSSVSFRSENRH 1039
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1041 SNTQGVVRCHLGOLAKGTVEVSGLLRVHNEFFRAKFKSLTVVSTFELGEE 1100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1040 RTASCNTMCMURDLQVKGFTLVNSTRIWNGTFASTFOYVQLTAAEIDIT 1099
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1101 TEASRWSLSLEVVQTRP---ILISLWILIGSVLGLLALLVFCMLKGFPA 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1100 EE---NVTYIPLTIMKHEKEVPTGVIVSGVIGILLALLVIMKLGFFKRYEM 1155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 -KIPEEKREKL 1165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1156 AKNPDETDTETEL 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

S44142

VLA-2 protein homolog - mouse

C.Species: Mus musculus (house mouse)

C.Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999

C.Accession: S44142

R.Jedelman, J.M.; Chan, B.M.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Danjanovich, L.; Latze

submitted to the EMBL Data Library, January 1994

A.Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not

A.Reference number: S44142

A.Accession: S44142

A.Status: preliminary

A.Molecule type: mRNA

A.Residues: 1-1178 <EDB>

A.Cross-references: EMBL:229987; NID:9473098; PIDN:CAAB2877.1; PID:9473099

C.Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

F.169-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 28.6%; Score 1744; DB 2; Length 1178;

Best Local Similarity 35.3%; Pred. No. 4.9e-122;

Matches 420; Conservative 221; Mismatches 479; Indels 70; Gaps 31;

```

9  LFLPLVFLTLGL---GSPNLDEHRLFPPEPEAFGYSVLQHVGGGGRMMVLGAPWDGP 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
10  LLLQLMLVQGLNLCLANVGLPKAKITSGSSSEQGYVQQLTNPPQGMMLLVSSPM 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
66  SGDRGDVYRCPCVGAHANAPCAKGH- GDYOLGNSHPAVNMHLSLLETDDGGGF 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70  PENRMGDVYKCPV-DLPATCEKLNQNASASISNTEITKTNLSGLTLTRNPGTG 128
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125  APLMRACGSSVFSSGICARVDASFPQGS LAPTAQRCPPTMDVIVLDSGNSIY 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129  GPLMAQCGNOYVATGICSDVSPDFQLTFSFPAVQACPLDVVVVVCDSNSIYPM 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
948  QTFRLRVKGLFIDPEQIOVGLVQGESPVHEWSLQDFRTKEVVAANKLSRERETK 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181  KNFLKLVKPTGLDYGKTKQVALIQYANPRRIIINLNDETEDVQALSETRQHG 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245  TQOAIIVACTEGFSQSHGGRPEARLLVVTTDESHDEGLPALKACAGATRYGIAY 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 249 TFRALFADVAVYSGTSGRPGATKVVVVVTDGESHGSKLTVIQCNDEILRFGLAV 308
 QY 305 LGHYLRQRPDPSSFLREITRTIASDPERFFNVNTDEALTDIVDALGRIFFLEGSHAEN 364
 Db 309 LGYLRNADLTNKLKEIKALSTPERKFRFVNADEALLEAKAGTLGRIEIEGT-VQG 367
 QY 365 ESSFGLMSQIGPSTHRL--KDGILFGVAVDVGWSVLMLEGHR--LFFPRMALEDEF 420
 Db 366 GDNFGEMNAQVGFSAVAFQNDILMLGAVGAFDMGSLIV-QETSHKPVIFP-----KQAF 421
 QY 421 PRALQ--NHAAYLGVSYSMLRGRRRLFLSGAPFRHRGKVIATQKKDGAIVRAQSIQ 478
 Db 422 DOVLDRNHSFLGSAVAISTEDGVH-FVAGAPPANVYQIVLVSVKQGVNTVYIQSR 480
 QY 479 GEOIGSFSELCPLDPTDGTDTTVLVLAAPMFLGPONKETRNVVYVVGQOSLLTLQOT 538
 Db 481 GQOIGSFSSVLCSDVDVDKDTITDVLVGAAPFYMDLKKBEKVLVFTTKILMLQHQFL 540
 QY 539 LQPEPPDARFGFANGALPDMLNODGFADYVAGAPLEDGHQALVLYHTQSGVPRHPAQR 598
 Db 541 EGPEGTGNARFGSALIALSDINMDGFNDVYVGSVVENENSGAVIYNGHQITRTKYSQK 600
 QY 559 IAAA--SNPHALSYGRSDVGRDLDDGDLVYVANGAQAALLSSRPVHLTPELEVP 656
 Db 601 ILGSGAFRRRLQGFGRSLDYGDLNGDSITVSGALGOVQLWMSOSIADVAIALFLTP 660
 QY 657 QAIYVQDRCRRGQEAVALTAALCFQVTSRTPGMDHQFYVRFASIDEMTAGARAADF 716
 Db 661 DKITILNDK-----ITKLCPRAFRPAGQ--NNQVALLFMNTLDADGHSRRVSR 711
 QY 717 GSGQRLSPRRLR--LSVGNV--TCEQLHFRVLDTSYLRVALVTYFALDNTTKPG--PVL 771
 Db 712 GVFRNSERFLQKNNVNVNEVKCSSEHHSIQKPSDVVNPDLDRVIDSEN--PQTSPL 768
 QY 772 NEGSTSTOKLVNPFKDCGPDNECTDVLQYNNMIRSKRAPPVYGRGRKRVLVSTLE 831
 Db 769 EAVSETVAVKFSIPFKECGSDIGCISDLILDV-QQLPALQTSFIVSNQNKRLTFPSVLK 827
 QY 832 NRKENAVNTSLIIFSRNLHLASLTPQRESPI--KVECAAPSARHL--CSVGHVPFQNG 887
 Db 828 NRGESANTVLAERSENLFPSFS---MPVDGIEVTCEVSSQKSVTCDDGVYALISE 883
 QY 888 AKVTFLEEFSSCSILSQVFGKLTASDSLERNGTLOENTQTSAYIOYEPHLLFSSSES 947
 Db 884 QCVETFINDFMLQNLQQAALNFQAFSESQETNNA--DNSVSLTPIPLYADEMLTST 941
 QY 948 TLHREVPHYGLP-----VGPPEPKTLFVQNLGCVVSGLIISALLPAVANGNYF 1001
 Db 942 NINFEISSDENAPSVIKSVEDIGPKFIPLKLV--TAGSAPVSMALVTIHPQYTERKQPL 1000
 QY 1002 LSLGQVITNNA---SCI--VQNLTEP---GCPVHPEBLQHTNRLNGSNVQGVVRCHLG 1053
 Db 1001 LYLTEIGIDQAGDICTAEINPLKLPHTAPSVSPKXNENFRHKEJEDCRTTSCSNITCWK 1060
 QY 1054 QLAKGTEVSGLLRLVHNEFFRRAKFKSLTVVSTELGTEGSSVLQLTASRWSLSLEV 1113
 Db 1061 DLHMAKEFIVNTVYVNRFTAFSTFQYQLFAAIEDITHNQ--LFVIEENAVTIPIMIM 1119
 QY 1114 VQTRILISLWLLIGSVGLGILLALLVFCMLKGLFF--AHKTI---PBE 1158
 Db 1120 KTEKAEPVPTGISIIAGILLLMTAGLWKLGFFKQYKQKMGQNPDE 1169
 RESULT 6
 S03308
 Cell surface glycoprotein CD11a precursor - human
 N:Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function
 C:Species: Homo sapiens (man)
 C:Date: 28-Feb-1990 #sequence revision 28-Feb-1990 #ext_change 20-Aug-1999
 C:Accession: S03308; A47565; A48759; S3604
 R:Larson, R.S.; Corbi, A.L.; Berman, L.; Springer, T.
 J:Cell Biol. 108, 703-712, 1989
 A:Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit

A:Reference number: S03308; MUID:89139587; PMID:2537322
 A:Accession: S03308
 A:Molecule type: mRNA
 A:Residues: 1-1170 <LAR>
 A:Cross-references: EMBL:Y00796; NID:931421; PIDN:CAA6747.1; PID:931422
 A>Note: part of this sequence was confirmed by protein sequencing
 R:Corneill, R.D.; Gollahan, K.A.; Hirstein, D.D.
 Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993
 A:Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) pro
 A:Reference number: A47458; MUID:93248261; PMID:8097887
 A:Accession: A47458
 A:Molecule type: DNA
 A:Residues: 1-20 <COR>
 A>Note: sequence extracted from NCBI backbone (NCBI:130862, NCBI:130863)
 R:Shelley, C.S.; Farokhzad, O.C.; Arnaout, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993
 A:Title: Identification of cell-specific and developmentally regulated nuclear factors t
 A:Reference number: A47565; MUID:93281759; PMID:8099450
 A:Accession: A47565
 A:Molecule type: DNA
 A:Residues: 1-20 <SHE>
 A:Cross-references: GB:M5609
 R:Nueda, A.; Lopez-Cabrera, M.; Vara, A.; Corbi, A.L.
 J. Biol. Chem. 268, 19305-19311, 1993
 A:Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.
 A:Reference number: A48759; MUID:93374910; PMID:8103515
 A:Accession: A48759
 A:Molecule type: DNA
 A:Residues: 1-20 <NUE>
 A:Cross-references: EMBL:Z22804; NID:9311405; PIDN:CAA80461.1; PID:9311406
 A:Gene: GDB:ITGAL; CD11A
 A:Cross-references: GDB:119757; OMIM:153370
 A:Map position: 16p11.2-16p11.2
 C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
 C:Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <
 F:154-317/Domain: von Willebrand factor type A repeat homology <VMA2>
 Query Match 18.1%; Score 1105; DB 2; Length 1170;
 Best Local Similarity 29.4%; Pred. No. 4.4e-74;
 Matches 366; Conservative 198; Mismatches 465; Indels 216; Gaps 55;
 QY 11 LPLVFLGL-----CSPNLDENHRLFPGRPA--EGYSVLQVHGCGQMWLVGAPWD 63
 Db 9 MAMLLSGFFFPADPASTYNDLVGRARSP--SPFRGRHFGYVLO-VGNG--VIVGAGE 63
 QY 64 GPGDRRGDYYRCVGAHNAFCAKGLGDYQLGNSHPAVNMHIGMSLETDDGCGFMA 123
 Db 64 GNS---TGLVQCGSGTCHCLPVT-----LQSNV--TSXYLGMTLADPTDGSILA 110
 QY 124 CAPLMSRACGSSVFPSSGICARVDASFO--PGGSLAPTYQRC--PTYMDVYIVLDSNSIYP 180
 Db 111 CDPGLSRCDQNTVYLSGICYFRQNLQGPMLQGRPGFOECIKGVNDVLFDFGMSLQPD 170
 QY 181 -MSEVQFLRLVAKLFLIDPEQIOVLQVQGESVHMSLGF--RTKEVVAAPAKLS 236
 Db 171 EFQKILDMKVMKKL--SNTSYQPAAVQFSTYKTEDESDYVKKMDPDLALGVKMKL 228
 QY 237 RREGREFTQAQIAIVACTEGFSQSHGRPEARLLVVVTDGSHGELPALAKACEAGR 296
 Db 229 LL-----TNTPAIVVATEVREELGARPDATKYLIIITDEALDSGNDIAKD----- 278
 QY 297 VTRYGIAVLGHYLRQRPSSFLREITRTIASDPERFFNVNTDEALTDIVDALGDRIFG 356
 Db 279 IIRYIIGIKRFPQKESQET-----LHKFASKPASEFVKLIDTPEKIKDLFTLEQKTIYV 333
 QY 357 LEGSHAENESSFGLEMSQIGPSTHRLKDLIFGVAVDVGWSVLMLEGHRLLPPRMAL 416
 Db 334 IEGTSKODLTSFNNELSSGISADLSRGHAYVAVGAKDMAGGLDKADLIQ----- 385
 QY 417 EDER-----PRALQNHAAVLYGVSYSMLRGRRRLFLSGAPFRHRGKVIATQKKDGA-V 471

```

Db      386 DDTFIGNELTPLEVRAGVLYTWTLPSSOKTSLASGAPRYQWGMRVLLFQEPQGGHM 445
Qy      472 RVAQSLQGEIQISYFSSSELCPDLTDGDTTDLVLAAPMFLGQNKETGRVYVY--LVG 528
Db      446 SOVQITIHGQIGSYFSGELCGVDVDDGETELLIPAFYEQ--RGRVFPYQRRQGG 503
Qy      529 QGSLTLQGLTLOEPQDARFPGMAGLPDLNODGFADVAVGAPEEDHQGLYLHGTQ 588
Db      504 FEVVSLSQG--DGYVL-GRFGAITALDINDGLVDVAVGAPLEE--QGAIVITNGNH 558
Qy      589 SGVRPAPQRIAAASMPHALSYFGSVGDLIDGDLVDVAVGAQAAIILSSRPYHL 648
Db      559 GGLSPQSPQRIEQTQYLSGIQWFGRSIHGVKDLLEGDLADVAVGAQSQIMVLSSREVDM 618
Qy      649 TBSLETPQALISVQVQDC---RRQGEAVCLTALCFQVNSTPBRMHQFMR----- 699
Db      619 VTLMSFSPAPLIPHEVECSYSTSNKKKEGVNIT--ICFOIKSLYP-----QFQGLVANI 671
Qy      700 -FTASLDEWTAGARAAPDGGQRLSPRLRLSVGNTCEQLHFNV-LDTSYLRPVALT 757
Db      672 TITLQDGHGRTRRGRGFPGRHNL--RNNIAVTSMSCIDFHFVQVQDLISPTNVSL 729
Qy      758 TVAL---DNTYKPGVNLNCSPTSIGL-----VPFSKDCGPDNECVTLVLQVN-MD 806
Db      730 NPSLMEEGETPRDQAGKDIPILRPSLHSETWEIPEFKNCGEDKCEANLRVSPSPAR 789
Qy      807 IGSRRAPFVNGGRKVLVSTLENKENAYNTSLIIFSRHLASLTPQR-ESPITY 865
Db      790 SRALRLTAPA-----SLSVLSLSNLEBDAYVQDLHPGGLSPRKVMKPHSQIV 843
Qy      866 ECAPASAHARL-----CSGVHPFOTGAKVTFLFEFSCSSLSQVFGKLTASS--DS 917
Db      844 SCEELPEERLLSRALSCVSSPIFKAGHSVA-----LQMMNTLVNSWMDG 891
Qy      918 LERNGT-----LQNTAQTSAVIOYEPHLF--SSESTLHRYEVHRYGTLFVPG 965
Db      892 VELHANVTGNEDSDLEDNSATTIIPILYPINILIQDEDSITLY-----VSFTKGP 944
Qy      966 PEKTLRLQNLGCVYVSGLIISALLPAVAHGNVYLSLQVITNNASCIQNLTPRP-- 1023
Db      945 -----KIHOVKMYOVRIPSI-HDN-----IPTLEA---VVGVPQPS 981
Qy      1024 GP-----PVHPEELQHTRLNGSNTQC---QVVRCHLQGLAKGTEVSGILR 1067
Db      982 GPTTHQMSVQMEPRVYCHEDLE--RLDPAERCPGLFRCPVFRQELIVQVIGTLE 1038
Qy      1068 LV---HNEFPRAKFKSLTVVST--PELGTEEGSVLQLTASRMSSELLEYVQTRPILI 1121
Db      1039 LVGELEASMFSLCSLSISFNSSKHFHLYGSNVSLAQVY-----MKVDVYVEKOMLY 1091
Qy      1122 SLWILIGSVLGGILLALLLVFCMLKGFPAHKKIPEEKREKLE 1166
Db      1092 -LVVLSG--IGLLLLLIIFVLYKGVFP-----KRNLEKME 1126

```

RESULT 7

156126 lymphocyte function-associated molecule-1-alpha - mouse

C:Species: Mus musculus (house mouse)

C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999

C:Accession: 156126

R:Kaufmann, Y.; Tseng, E.; Springer, T.A.

J. Immunol. 147, 369-374, 1991

A:Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit

A:Reference number: 156126; PMID:91268576; PMID:2051027

A:Accession: 156126

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1163 <RES>

A:Cross-references: GB:M60778; NID:q198785; PIDN:AAA39426.1; PID:q198786

C:Genetics:

A:Gene: LFA-1

```

C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
F:151-315/Domain: von Willebrand factor type A repeat homology <WMA1>
Query Match      18.0%; Score 1097; DB 2; Length 1163;
Best Local Similarity 28.4%; Fred. No. 1.7e-73;
Matches 356; Conservative 192; Mismatches 470; Indels 234; Gaps 48;
Qy      12 PLVFTGTG-----CSFPLNDEHHPLFPPEAEFGYVLQVHGCGQRMVLGAPWDPDS 66
Db      8 PRLLILGLQLAKAMSYVLDTRPQSLAQAQRHGHQVQLDEG---VVVGAPGE--- 60
Qy      67 GDRRDVYRCVPGAGAHNAPCAKHLGDYQLGNSSHPAVNMHLGSLTDDGGFMACAP 126
Db      61 GDNLTGLYHCRTPSSEFCQVPS-----LHGSNH--TSKYLGMTLATDAKAGSLIADCP 110
Qy      127 LMSRAGSSVSSGICANVADSFOQSGSLATPQRCPTV-----MVVVLIDGNSNI- 178
Db      111 GLSRTCDONTYLSGCL---YLFPQSLGEPMLQRPAYQCMKGVLDVLPDSSQSLD 165
Qy      179 -YPMGEVOTFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSLGFRTKEEVRAKANSR 237
Db      166 RKDFEKILFEMKDVNRKL--SNTSYQFAVQFSTDRCRETFEFDY-----VKQKNPDV 217
Qy      238 REGRE-----TKTAQAIMVACTEGSFQSQHGRPREARLLVVTVDGESHGDELPALKA 291
Db      218 LLGSVQPMFLTLNTRFRAINYVAHVFKEESGARPDATKVLITDGEASDKGNISA--- 273
Qy      292 CEAGVTEYGVAVLGHVYRQRDPSSFLREITIASDDEEFNFVNTDEALTLTDVALG 351
Db      274 -ADHTYITIGIGHFVSVQKQ---KTHIFPSEVEEVEVKILDTFEKIKDLFTDLO 326
Qy      352 DRIFLESASHAENESSFGLEMSQIGFSTHRLKDLGLFGMVGAYDMGGSVLMEGHRLEP 411
Db      327 RRIYIBETNQDLTSFMELSSSGISADLSKHAHVAVGAKMAGGFLDLR----- 379
Qy      412 PRMALED-----EPPPALQNHAAVYIGYVSSMLRGLRRLISGAPRRRHKKVIAF 463
Db      380 -----EDLQATFVQGEELTSDVREGYLYTVAMWTSRSSRPLLAAGAPRYOHQVQLVF 434
Qy      464 QL-KKDGAVRAAQLQGEIQISYFSSSELCPDLTDGDTTDLVLAAPMFLPQNKETGRV 522
Db      435 QAPPEGKMNQOTOKIEGTQISYFSGELCSVDLDQGAELLLGAPLPFQEQ--RGRV 492
Qy      523 VYVILQGQSLTLTQCTLOPEPPD--ARFGMAGLPDLNODGFADVAVGAPEEDHQGL 581
Db      493 FTY-QRRSLTEWSELQDGPYGLRGGAITALTIDINGRDLVDVAVGAPLEE--QCAV 549
Qy      582 YLVHGTQGVRRPAPQRIAAASMPHALSYFGSVGDLIDGDLVDVAVGAQAAIILLS 641
Db      550 YIFNGKPGGLSPQSORIQGAQVPGIRFWFGRSIGHVLDLGDRLADVAVGAEGVVVLS 609
Qy      642 SRPIHLPSLEVPQALISVQVQDC---RRQGEAVCLTALCFQVNSTPBRMHQF 697
Db      610 SRPVVDVTELSFSPEELPHEVECSYAREQKHGVLYKA--CFRIKPLTP-----OFQ 662
Qy      698 MR-----FTASLDEWTAGARAAPDGGQRLSPRLRLSVGNV-----TCEQLHFNV-L 744
Db      663 GRLLANLSYTIQLODHRMRSGLPDQSGHELS-----GNTSITPDKSLCDFHFRPI 714
Qy      745 DTSYLRPVALTVPFALDNTTKPGVNLNCSPTS-----SIQKLVPSKCD 788
Db      715 CIQDILSPINSLNLSL-----LEEEGTPRQOKGRAMOPILRPSIHTYKELPFPEKN 766
Qy      789 CGPDNECVTLVLQVNMDIRGSRKAPFVVRGGRKVLVSTLENKENAYNTSLIISFR 848
Db      767 CGEDKKCEANLTLS---SPARSGPLRMS--ASLAEMTLLNSGEDAVVWRDLDPFR 820
Qy      849 NLHLASLTP-QRESPIKECAAPSAHARL-----CSGVHPFOTGAKVTFLLEFEFSCS 901
Db      821 GLSPKVMQLQPHSMRPSCEELTEGSLTKTLKCNVSSPIFKAGQSVS----- 870
Qy      902 SLLSQVFGKLTASS--DSLERNGTL-----QENTAQTSAVIOYEPHLFSSSESTL 949

```


Db 871 --LQWENTLNSNDEPELNGTACENENSLQEDNSAARIIPVLYNITLLEQEN- 927
 Qy 950 HREYVHPYGTLPVGPGEFFKTLRVONT-----GCYVSGLIISALLPAVAHGQVPL 1002
 Db 928 -----STLYISFPTKPKPTQOVHVVYVRIQPSAVDHNNPTLEALGVPRHSEDDI 979
 Qy 1003 SLSQVITNN--ASCIYQNTTEPPGPPVHPEELQHNRLNGSVTQCVV-RCHLGQLAKT 1059
 Db 980 TTTWVGQDPLVTCHSEDDIKRPSSEABQP-----CLPGVQRCPIYFRWEILLQVGT 1032
 Qy 1060 EVSVGLRLVHNEFPRAKFKS-LTVVSTFELTEGEGVLTQLEASRMSES----LLEVV 1114
 Db 1033 V-----ELSKETIASSTLSICSLSVSNSSKRFHL-YGSKASEAVLVKVDLI 1080
 Qy 1115 QTRPILISMLIGSVLGLLALLVFLCMKLGFFAHKKTPEEKREKLE 1166
 Db 1081 HEKENTL-HYVVLVSG--IGGLVLLPLFLALYKVGFF-----KNNLKEKME 1122
 RESULT 8
 RHWUB
 cell surface glycoprotein CD11b precursor [validated] - human
 N.Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Mac-1
 C.Species: Homo sapiens (man)
 C.Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text change 31-Dec-2000
 C.Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567
 J.Corbil, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.
 J. Biol. Chem. 263, 12403-12411, 1988
 A.Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD11b)
 A.Reference number: A31108; MUID:88315033; PMID:2457584
 A.Accession: A31108
 A.Molecule type: mRNA
 A.Residues: 1-1153 <COR>
 A.Cross-references: GB:J03925; NID:9187284; PIDN:AAA59544.1; PID:9307148
 A.Note: part of this sequence was confirmed by protein sequencing
 R.Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.
 J. Cell Biol. 106, 2153-2158, 1988
 A.Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor Mac-1
 A.Reference number: A28915; MUID:88257215; PMID:2454931
 A.Accession: A28915
 A.Molecule type: mRNA
 A.Residues: 1-959, 501-965, 'P', 967-1153 <ARN>
 A.Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:9186935; PIDN:AAA594
 A.Note: the authors translated the codon TAG for residue 1129 as Thr
 R.Shelley, C.S.; Arnaout, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991
 A.Title: The promoter of the CD11b gene directs myeloid-specific and developmentally regulated expression
 A.Reference number: A41600; MUID:92073318; PMID:1683702
 A.Accession: A41600
 A.Molecule type: DNA
 A.Residues: 1-9 <SHR>
 A.Cross-references: GB:M6724; NID:9180018; PIDN:AAA58410.1; PID:9553215
 R.Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.
 Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988
 A.Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesion receptor
 A.Reference number: A41933; MUID:88190151; PMID:2833753
 A.Accession: A30892
 A.Molecule type: mRNA
 A.Residues: 917-1042 <AR2>
 A.Cross-references: GB:M18044
 R.Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.
 Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989
 A.Title: cDNA sequence of the alpha subunit of the human neutrophil adherence receptor
 A.Reference number: A32218; MUID:89098933; PMID:2563162
 A.Accession: A32218
 A.Molecule type: mRNA
 A.Residues: 9-1153 <HIC>
 A.Cross-references: GB:J04145; NID:9189068; PIDN:AAA59903.1; PID:9386975
 A.Note: part of this sequence was confirmed by protein sequencing
 R.Fleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.

J. Immunol. 150, 480-490, 1993
 A>Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1 during evolution.
 A.Reference number: A46526; MUID:93123748; PMID:8419480
 A.Accession: A46526
 A>Status: not compared with conceptual translation
 A.Molecule type: DNA
 A.Residues: 1-499, 501-1153 <FILE>
 A.Cross-references: GB:S52227; NID:9263047; PIDN:AA824821.1; PID:9263049
 A.Note: the last three bases of intron 13, CAG, are included in some but not all mature transcripts
 R.Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.
 Biochem. Biophys. Acta 874, 368-371, 1986
 A>Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across species
 A.Reference number: A90664; MUID:87076671; PMID:3539202
 A.Accession: A26091
 A.Molecule type: protein
 A.Residues: 17-31 <PIB>
 A.Experimental source: granulocytes
 R.Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.
 Blood 79, 865-870, 1992
 A>Title: Characterization of the myeloid-specific CD11b promoter.
 A.Reference number: 152567; MUID:92144986; PMID:1346576
 A.Accession: 152567
 A>Status: translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-9 <RES>
 A.Cross-references: GB:M84477; NID:9180184; PIDN:AAA51960.1; PID:9553219
 C.Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1
 C.Genetics:
 A.Gene: GDB:ITGAM; CR3A
 A.Cross-references: GDB:120599; OMIM:120980
 A.Map position: 16p11.2-16p11.2
 A.Note: promoter contains a GATA motif and two Sp1 consensus binding sites
 C.Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
 C.Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag
 F.11-16/Domain: signal sequence #status predicted <SIG>
 F.17-1153/Product: cell surface glycoprotein CD11b #status experimental <MAT>
 F.17-1108/Domain: extracellular #status predicted <EXT>
 F.148-318/Domain: von Willebrand factor type A repeat homology <WMA2>
 F.445-473/Region: calcium/magnesium binding #status predicted
 F.530-538/Region: calcium/magnesium binding #status predicted
 F.593-601/Region: calcium/magnesium binding #status predicted
 F.1109-1134/Domain: transmembrane #status predicted <TM>
 F.1135-1153/Domain: intracellular #status predicted <INT>
 F.86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding

Query Match 17.8%; Score 1085.5; DB 1; Length 1153;
 Best Local Similarity 29.6%; Pred. No. 1.2e-72;
 Matches 366; Conservative 196; Mismatches 492; Indels 183; Gaps 47;

Qy 11 LPVFLTC--LCSPFNIDENHRLPFGPREAREGVSYLVQHGCGORMMLVGAPMDP89GD 68
 Db 3 LRVLLTALTLTCHGFNDLTENAMTFQENARG-FQGSVQVQSGR--VVGAPOETIVAN 58
 Qy 69 RRGVYRCPPVGAGNAPCAKGLHDYQVQNSHP-----AVNMHGMVLETDGPG 120
 Db 59 QRGSLYQC-----DYSTG-SCEIRLOVPEAVMMSIGLSLAATSPQ 101
 Qy 121 FMACAPLWSRACGSSVSSGICARVDSFQPGSLAPTAOR-CPYV-MDVVIVLDGNSI 178
 Db 102 LLACGPVHQCSNRYVKGICPLFGSNLRQPPQKFPBALRGCPQEDBDIFLDGSGSI 161
 Qy 179 YPMSEVQTLRLVGLKFLID-FQIQ-----VALVOGESPVHMSIGDRTAEVVRA 231
 Db 162 IPHD-----FPRM--KEVSTVMQQLKSKTLTFLIMQVSEFRITFTKEQNNPNPSL 214
 Qy 232 AKNLSREGRETKTAQAIWVACTGFSQSHGREGREARLIVVVDGSSHDEEPALKA 291
 Db 215 VKPITQLGR-THATGIRKVRVRELFNITGARNAKILVITDGKEF-GDPLGYEDVI 272
 Qy 292 CEAGR--VTRYGLAVLGHVLRQRPDPSSFLREITIASDPDERFFVNTDBAALTDI 349

```

Db      273 PEADREGVIRYIGV-GDAFRSEKS-----ROELINTIASKPRPDHVFQVNNFALKTIONQ 327
Qy      350 LGDRIFGLGSHAENESSFGLMSOIGFSTHRLKOGILFGMGAYMGWSVLMLGEGHRL 409
Db      328 LNEKFAITGTOTGSSSEHEHESQGFSAITSNPPLSTVGSIDMAGCVFLYTSKXS 387
Qy      410 PPRPMLEDEFPFPAONHAAYLYGSVSMILRGRRLLFSGAPRRFRHKVIAFOLKKG 469
Db      388 TFINTRVDS-----DMNDAYIGYA-AAILNRNVGSLVGAAPRGHILVAMFR-QNTG 440
Qy      470 AVRVAOSLOGEOIGSFSELCPLDTRDGTDDVLLVAPMELGPNKTEGKYVYLV-- 527
Db      441 WMESNANVKGTOIGAFGASLCSVDNSGSDTLVIGAPHYY--EOTRGGVSVCPLE 498
Qy      528 GOOSLLTLTGLOPEPPOD--ARFGFAMGALPDLNOCGFADVAVGALPEDHGOALYLYG 586
Db      499 GGRARKOCDAVLXGEGQGWGRFGAALTVLGDVNGDKLTDVAIGABGEDNDGAYVLFPG 558
Qy      587 TG-SGVPRPAPORIAAASMPHALSYFGRSVDGRLDGDLDLVAVAGAAIILSSRPI 645
Db      559 TSGSGISPSHSQRIAGSKLSPRLQYRGSLSGQDITMGVLDITVGAQHVLILRSQPV 618
Qy      646 VHLTSLSEVTPQAIISVQDRCR--RQGEAVCLTALCFQYTSRTPRGWDH--QFTWR 699
Db      619 LRVKAIMENPREVANVEECNDQVVKKEAG--EYRVCLVYQKSTRRLREGOIQSVVT 676
Qy      700 PFAASLDEMTAGARAFAFDGSGRLSPRLLSVG--NTVCQELPHVLD--TSDVLRPALTV 757
Db      677 YLALDLSGRPHRAVAFNET--KNSTRQTOVIGLTOCTTLKQLPNCIEDVSPVLRL 734
Qy      758 TFLALNTTKPG-----PVINEGSPSIOKLVPESKDCGDNCEVTDVLQ--VNMDIRG 809
Db      735 NSLSVGTPLSAPGNLRPLVAEDAQRLETLFPEKKGNDNICQDDLSITFSFMSDC-- 792
Qy      810 SKKAPVAVRGGRKKVLVSTTLNKRKNANVTLSISIFSKNL--HLASTPQR--ESPILV 865
Db      793 -----LVVGGRFEFNVTVTVRNDGDSYRTQVTFEPFLDLSYRKVSTLQNRORSQSMRL 846
Qy      866 ECAPSA-----HARLCSGVHPVQTCAGKVTFLLEFFSCSSLSQVFGKLTASSDRL 918
Db      847 ACESASSTSVGALSTSCSINHPIPENSEVTFNITFVDSKASIG--NKL-----L 898
Qy      919 ERNGTLQENTAOJ-----SAYIOYEPH-----LLFSESTLHREYVHPYGT 959
Db      899 KANVTSSENMMPRTNKTPEQLLQVKAAYVMVTVSHGVSTKYLNFATSEVTSVMQOY-- 956
Qy      960 LPEVGPEPEKTLRQNLGCIYVSGILISA--LLPVAHAGNYFLSLSYT----- 1008
Db      957 -----QVSNLG--QRLPLISLVLP-----VLNQTVMWRPOVTF 991
Qy      1009 TNNASCIQNLTEPRGPVPVHPBELQHTNRLNGSNTOCVVVRCHLQGLAGTEVSGVLR 1068
Db      992 SENLSTCTKTERLBSHSPFLAEKRAPVNGSIAVCORIQCDIPFGIOEFPNATLKN 1051
Qy      1069 VANEFFRRAKFKSLTVVSTFELGTEGVLQLEASRWSESLLEV-VQTRPILISLWLI 1127
Db      1052 LSFDMYIKTSHNHLITVSTAILFNDVSFTLLPGQCAFVRSQETKEVBEFVBNPLPLIV 1111
Qy      1128 GSVLGLLLALLVFLMKLGFPAHAKKIEEEKREK 1164
Db      1112 GSSVGLLLALLVITVLYKLGFF-----KROYK 1139

```

RESULT 9

cell surface glycoprotein CD11c precursor - human
 N:Alternate names: leukocyte adhesion receptor p150.95 alpha chain
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence revision 31-Dec-1992 #text_change 22-Jun-1999
 C:Accession: A36584; A35543; S00864
 R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
 J. Biol. Chem. 265, 12750-12751, 1990
 A:Reference number: A36584

```

A:Contents: erratum
A:Accession: A36584
A:Molecule type: DNA
A:Residues: 1-1163 <COR>
A>Note: this revision to the sequence from reference A35543 includes the carboxyl end
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.
J. Biol. Chem. 265, 2782-2788, 1990
A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150.95 molecule.
A:Reference number: A35543; MUID:90153906; PMID:2303426
A:Accession: A35543
A:Molecule type: DNA
A:Residues: 1-834 <CO2>
A>Note: this sequence has been revised in reference A36584
R:Corbi, A.L.; Miller, L.J.; O'Connor, K.; Larson, R.S.; Springer, T.A.
EMBO J. 6, 4023-4028, 1987
A:Title: cDNA cloning and complete primary structure of the alpha subunit of a leukocyte
A:Reference number: S00864; MUID:86166645; PMID:3327687
A:Accession: S00864
A:Molecule type: mRNA
A:Residues: 1-755, 'L', 757-1163 <CO3>
A:Cross-references: GB:M81695; EMBL:Y00093; NID:9487829; PIDN:AA59180.1; PID:9487830
A>Note: part of this sequence was confirmed by protein sequencing
C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my
C:Genetics:
A:Gene: GDB:ITGAX; CD11C
A:Cross-references: GDB:119758; OMIM:151510
A:Map position: 16p11.2-16p11.2
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;
F:1-19/DNA: signal sequence #status predicted <SIG>
F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MNT>
F:120-1107/DNA: extracellular #status predicted <EXT>
F:1149-319/DNA: von Willebrand factor type A repeat homology <VWA4>
F:1108-1133/DNA: transmembrane #status predicted <TM>
F:1134-1163/DNA: intracellular #status predicted <INT>
F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (asn) (covalent) #status pre

```

Query Match 17.3%; Score 1054; DB 1; Length 1163;
 Best Local Similarity 28.5%; Pred. No. 2.9e-70;
 Matches 356; Conservative 201; Mismatches 474; Indels 220; Gaps 50;

```

Qy      13 LVFLTGCLSP--FNIDEHHPRLFPGEPAEFGVYLQHVGGQRMVLGAPWDCGRCRR 70
Db      8 LLLFTALATSLGFINDTLELTAFL-RVDSAGCGDSVQVYANS---WVVVGAPOKITAANQT 63
Qy      71 GDVYRCFPGAGAHNAPCAAGHLDGVLQNSHP--AVNMHLGMSLLETDDGPFMACAPLMS 129
Db      64 GGLYQC--GYSTGACE-----PIGLQVPEAVNMSLGLSLASTISPSQLACGPVH 113
Qy      130 RACGSSVPSGICARVADSFOGSLAFTAORCPY--MDVYIVLDGNSNIYP--WSEVQT 186
Db      114 HECGNMVLTLGCLFLGPT-QLTQRLPVSRQECRQEDVIFLLDGSISSRNPATMWN 172
Qy      187 FLRLVGLFTIDPEQIVGLVQGESPVHEWSLGFRTKEEVVRAAKLSRREGRETITA 246
Db      173 FVRAVISO-FQRP-STQPSLMQFSNKQFHTFTEFFRTSNPLSLASVHQLQG--FTYTA 229
Qy      247 QAIWACTGESSGSGRPEARLLVVVTDGESH-DGEELPAALKACAGVTRGIVLV 305
Db      230 TAIQNVHRLPHASGARDATKILIVITDGKKGDSLDYQDVIPMADAGIRIYALVG 289
Qy      306 GHYLRQRDPSSFLREITIASDPPERFFVNTDEALTDIVDALGRIIFGLEGSHAENE 365
Db      290 LAFQNR-----NSMKELNDISKSQEHIFVDEPDAKDIONQKEXIFAIETETSS 344
Qy      366 SSFGLMSQIGPSTHRLKOGILFGMGAYDWGGSVLMLEGHRLPPPMALDEDEPPALQ 425
Db      345 SSFELEMAQEGFSAVFTPDGCVLGAVSGFTW-----SGAFLYPPNMS-----PTFI 391
Qy      426 NHA-----AYIGYSVSMILRGRRLLFSGAPRRFRHKVIAF-OLKKGAVRVAQS 476
Db      392 NMSQENVMDRSDYLGYSITELALMKGVQSLV-L-GAPRYOHTGKAVIFTOVSROW--RMKAE 448

```

```

QY 477 LQGEQIGSYFSGELCPDLTDGCTDVLVAAPMGLGPNKKTGRVYV----- 524
D 449 VTGTIGSYFYGASLCSVDVDTGSDTDLVIGAPHYV--EQTRGGQVSVCPPLPRGRMRWC 506
QY 525 --LVVGGQSLTLTLOGLPEPPDARFPGFAMGALPDLDNDGADVAVAGPLEDHOGALY 582
D 507 DAVLVEGQG-----HPW-----GRFGALTVLGVGVNDKLTLDVIGAPGEENRGAVY 554
QY 583 LVHGQTQS--GVRRPAPORIAAAMPALSYFGRSVDRGLDLDGDDVDVAVAGQAAILLS 641
D 555 LHHGVLPSPISPSHRSRIAGSGLSSRLQYFGQALSGGDLTDGDLVAVAGRGVLLIR 614
QY 642 SRPIYHLTPSLLEVTPAISVQVDCRR--GQAVCLTALCFQVTSRTP--GRMDHQFY 697
D 615 TRPVLMVGVSMQFIAPAEIPRSAPFECREQVSEQLTVQSNICLYIDKRSKNLGRDLQSS 674
QY 698 MFFTASLDWTAGAAPAFDGGQRLSPRLRLSVGNVTCQQLHFVLDTS--DYLPRPA 754
D 675 VTLDLALDGRSLSPRTFQETKNR--SLSRVRLGLKANE--NFMLLPSCYEDSVTPIT 731
QY 755 LVTFTALDNTTKP-----GPVLNKGSPTSIOKLVPFGKDCGPNCEVTDVLQVNDI 807
D 732 LALNFTL--VGKPLAFRLRLPMLAADQRYFTASLPFEKNGGADHICQDNIGISFPF- 788
QY 808 RGSRAKPFVVRGRRKRVLTLE-----NRKNAVNTSLSIIFS----- 847
D 789 -----GLKSLVGSNLELNAEVMVNDGDSYGT--TTFSPAGLSYRYVAE 834
QY 848 -----RNLHLASLTPQESPIKVECAAPSAHAALCSGHPVQOTGANVTLLFEPSC 900
D 835 GQKQGLSLHLTC---DSAPV---GSGGTWSTSCRINHLIFGGAQITFLATFVDS 886
QY 901 SSLLSQVFGKLTASSDSLEKNGTLOENTAO-----TSAYIQEPEHLFS--- 944
D 887 KAVLDRL--LTLANVSSENNTRTSKTTQQLPVKYAYTVVSSHQETKTLNSESSE 945
QY 945 --SESTLHRYEVHPYG--TLPVGPGEFETTLRVONLGYVSGIILSAL--PAVAHG 998
D 946 KESHVAMHRYGVNNIGORDLPV-----SINFVPEVLELQEAVMMDVEVSHQ 992
QY 999 NYFSLSQVITMNACTIVONLTERPGRPVHPELQHTNR--LNSNTQCVVRRCHLQ 1055
D 993 NPSLCS-----SEKLAIPA--SDPLAHIQNPVLDSIACQLRFRCVDSF 1037
QY 1056 AKGTESVGLRLVHNEFFRAKFKSLTYVSTFELTEBGSVLQTEASRMSSESLLEVQ 1115
D 1038 SVQELDPLTKNLSFGWKRLQKKVSVSAEITFDTSVSQLPGQAFPRKRA----- 1092
QY 1116 TRPILISMI-----LIGSVLGGILLALLVFCIMKLGFF--AHKQIPPE 1158
D 1093 TTVLEKVKVHNPTPLIVGSSIGILLALLITAVLYKVGFFRKQYKEMEE 1143

RESULT 10
S00551
Leukocyte surface glycoprotein Mac-1 alpha chain precursor - mouse
N:Alternate names: complement-3 receptor alpha chain
C:Species: Mus musculus (house mouse)
C:Date: 30-Sep-1989 #sequence-revision 30-Sep-1991 #text-change 22-Oct-1999
C:Accession: S00551; 159078
R:Pyteia, R.
EMBO J. 7, 1371-1378, 1988
A:Title: Amino acid sequence of the murine Mac-1 alpha chain reveals homology with the
A:Reference number: S00551; MUID:862812584; PMID:3044779
A:Accession: S00551
A:Molecule type: DNA
A:Residues: 1-1153 <PVT>
A:Cross-references: EMBL:X07640; NID:952982; PID:CA0479.1; PID:952983
A:Note: the authors translated the codon CAC for residue 569 as Gln
R:Sastre, L.; Roman, J.M.; Teplow, D.B.; Dreyer, W.J.; Gee, C.E.; Larson, R.S.; Roberts,
Proc. Natl. Acad. Sci. U.S.A. 83, 5644-5648, 1986
A:Title: A partial genomic DNA clone for the alpha subunit of the mouse complement recep
A:Reference number: I59078; MUID:86287312; PMID:2942940

```

```

A:Accession: I59078
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 11-44 <RES>
A:Cross-references: GB:M14293; NID:g198993; PID:AAA39484.1; PID:g554193
C:Genetics:
C:Gene: Mac-1
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C:Keywords: cell adhesion; glycoprotein; transmembrane protein
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-1153/Product: leukocyte surface glycoprotein Mac-1 alpha chain #status experimental
F:1148-318/Domain: von Willebrand factor type A repeat homology <VMA2>
F:1106-1129/Domain: transmembrane #status predicted <TM>

Query Match 16.7%; Score 1022.5; DB 2; Length 1153;
Best Local Similarity 28.6%; Pred. No. 6,66-68;
Matches 353; Conservative 218; Mismatches 481; Indels 181; Gaps 50;

QY 13 LVFTLGLCSFNLDEHHRLEPPGEAEFGYSVLQHWGGGRMLVGA PMDPSGDRRD 72
D 7 LVTLALCHGNLDTHEHMTFQENAKG-FGQNVVQ-LGTS--VVAAPQKAKAVNQGA 62
QY 73 VYRCPVGAHNAAPCAKHLGDYQLNNSHP-----AVNMHLGMSLLEFDGCGFMAC 124
D 63 LYQC-----DYST--SRCHPIPLQVPEPAVMSGLSLAVSTVPQQLAC 105
QY 125 APLMSRACSSVFSSGICARVDAS--FOPQSLAPTAQCPY--MDVVLVLSNSI--YP 180
D 106 GPYHONCKENTYVNGLCYLFSGNULRPPOQPEALBPCPOQESDLYFLVLSGSSINNID 165
QY 181 WSEVQFLRLVGLKFLIDPEQIQ-----VGLVQYGESFVHMSLGFRTKEEVRAAKNL 235
D 166 FQKMEFVSTW-----EQFKSKTLFSLMQYSEDFRIHFTNDFKRNPSRSHVSP 218
QY 236 SRREGRETKQALMVACTEFGSOSHGCRPAAALLVYVTDGESHDELPALAKACAG 295
D 219 KQLNGR--KTKASGARKVRELPHKTNGARENAAITLVITDGEK--GPDLYKQVIRPAD 276
QY 296 R--VTRVGIAYLGHYLRQRPSSFLREIRITASDPDERFPFNVTDEALTDIVDALGDR 353
D 277 RAGVIRYIVGY-GNAFNK---PQS--RELDITAKRPAHEHVFQVDFALMTIQNLQEK 331
QY 354 IFGLEGSHAENESSFGEEMSGQIGFSTRKLDGILFGWGVAYDM--GGSVLMLEGGRLEPP 412
D 332 IFALIEGTGTSTSSBEHMSQEGFSASITSNPILGSGVDMAGAFLYTSKQVPTIN 391
QY 413 RMALDEFPPLAKHNAAYLYGVSSSMLIRGGRPLFLSGAPFRFRKVIYALQKKQAVR 472
D 392 TTRVDSDM-----NDAYLGVA--SAVILRNVOGLVLGAPRYOHIGLVVMPR--ENFGTWE 443
QY 473 VAOSLQGEQIGSYFSGELCPDLTDGCTDVLVAAPMFLGPQKKTGRVYVLYGQ--- 529
D 444 PHTSKGQIGSYFSGASLCSVDMDADGNTNLTILGAPHYV--EKTRGGQVSVCPPLPRRA 501
QY 530 --QSLLTLOGLTLOPEPPDARFPGFAMGALPDLDNDGADVAVAGPLEDHOGALYLYHG 587
D 502 RMQGEALLHGP-QGHR--WGRFGAALTVLGVGVNDKLTLDVIGAPGEENRGAVYIFGA 558
QY 588 O-SGVRRPAPORIAAAMPALSYFGRSVDRGLDLDGDDVDVAVAGQAAILLSRPIV 646
D 559 SIASISASHSHRIIGAHSPGLQYFGQSLSGKDLTMDGLMDLVAVAGQHLLTLRAQV 618
QY 647 HLTPSLLEVTPAISVQVDCRR-----GQAVCLTALCFQVTSRTPR---WHDQF 696
D 619 RLEATMESPKKVARSPACQEVLYKNKADGEVYCL-----RVKNTKRLBEGDIQS 672
QY 697 YMRFTASLDWTAGAAPAFDGGQRLSPRLRLSVGNVTCQQLHFVLD--TSYLRLPYAL 755
D 673 TVTVDLALDVPVRSRIAPFDET--KNNTRRRQVGVLMQKCTKTLTLIPDCVDSVSPITL 731
QY 756 TVTVALDNTTKP-----GPVLNKGSPTSIOKLVPFSKDCGPNCEVTDVLQVNDIR 808
D 732 RLNYTL--VGBPLRSFGNLRVLAMDQRFPTAMFPEKNGCNDISICDD--LSITWSAM 787

```

```

OY GSRKAPFVRCGRKRLVLTSTLENNKENAVNTSLIIFSRNHL-----ASLTPORESPIK 864
Db 788 G---LDTLVGGPQDFNMSTVLRNDGEDSYGIVQVTVYPSGLSTRKDSASQNPILTKRWPF 844
OY 865 VECAPSA-----HARLCS-----VGHPIVFQGAKYTFLEFEF-SCSSLSLOVFGK-LTAS 914
Db 845 VKPAESSSSSECHGALKSTWNINHPITFPANSEVTFNVTFDVDSHA5FGNKLLIKAIVAS 904
OY 915 SDSLERNGLTOENTQOTSAVIOYEPHLLFSS-ESTL-----HREYENHPY 957
Db 905 ENNMSRT---HHTKFOLELPVKYALMYMTVTSDESSIRVLFNFTASEMTSKVTOHOQOFNML 961
OY 958 G--TLPGCPPEFKTTLRVONIGCVVSGLIITSLALPAVAGNFYLSLOVTNNASCI 1015
Db 962 GQRLSPV--SVYFWLPIVOQINNTVWDHPQVIFSONLSACH-----1000
OY 1016 VQNLTEPBGPPVPR---ELQHTNRKLNGSNTQOCVVCRHQLQAKTEVSAGLLRLVANE 1072
Db 1001 ----TEQKSPF-HSNFMRDQLERTPYLNC5AVACRKIQCDLPSFNQOELFNTTLKGNLSFD 1055
OY 1073 FPRRAKFKSLVAVSTFELGTEBG5VLQLTASRMS5SLLEV-VQTRPILISLMLIGSVL 1131
Db 1056 WYIKTSHGHLVLVSTTEILFPND5AFALLPGQESYVR5STETKRVBEYVHNPPVLIVGSSI 1115
OY 1132 GGLLTLALLVFCMLKMGFFFAHKKIPEEKREKX 1164
Db 1116 GGLVLLALITAGLYKLGFF-----KROK 1139

```

158409
 Integrin alpha-9 chain precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 31-Dec-2000
 C:Accession: I58409; A49459
 R:Hihi, K.; Yamakawa, K.; Ueda, R.; Horio, Y.; Murata, Y.; Tamari, M.; Uchida, K.; Takah
 Onogene 9, 611-619, 1994
 A:Title: Aberrant upregulation of a novel integrin alpha subunit gene at 3p21.3 in sma11
 A:Reference number: I58409; MUID:94119603; PMID:8290272
 A:Accession: I58409
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1035 <RES>
 A:Cross-references: GB:D25303; NID:9464180; PID:BAA04984.1; PID:9533327
 R:Palmer, E.L.; Ruegg, C.; Ferrando, R.; Pytel, R.; Sheppard, D.
 J. Cell Biol. 123, 1289-1297, 1993
 A:Title: Sequence and tissue distribution of the integrin alpha9 subunit, a novel patrine
 A:Reference number: A49459; MUID:94064789; PMID:8245132
 A:Accession: A49459
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 30-1035 <PAL>
 A:Cross-references: GB:L24158
 C:Superfamily: integrin alpha-4 chain
 C:Keywords: glycoprotein; metal binding; transmembrane protein
 F:1-27/Domain: signal sequence #status predicted <Sig>

Query Match 12.1%; Score 738; DB 2; Length 1035;
 Best Local Similarity 22.6%; Pred. No. 1.3e-46;
 Matches 284; Conservative 179; Mismatches 446; Indels 350; Gaps 44;

9 LFLPLVFLTGLCSFNLDEHNRPLFPGRPEAFEGSVLQHVGGGQRMVLGAPMD---G 64
 16 LLLALVVGAGIPAGAVNLDPQRFVHFGPADSFQVAVLEHFDNTRMVLVGA.PKADSKS 75
 65 PSQDRGDIYRCVPVGAANAP-----CAKGHLDYQLGNS-SHPAVNMLGMSL-E 114
 76 PSVKSPGAVFKCRV---HTNPDRCETELDMARKKRGTSCKTKCEDREDDEMVGSLARQ 132
 115 TDGDDGFMACAPLWBRAC--GSSVFSGICARVDASFPQGSGLATPACRPTVMVIVL 172
 133 PKADGRVLLCAKRMKITYEADHILPHGCIYIIPSLQAKG----- 173
 173 DGSNSIYPMSEVQTEFLRLVGLFTDPEQIQVGLVQYGESPVHEMSLQDFRKEEVRAA 232
 174 -----RT----- 175
 233 KNLRSREGRETKTAQALINVACTEGSQSHGSRPEARLLVVTDESHDEELPALKAC 292
 176 -----LIPCYEYKKY-----GEEHG-----SC 194
 293 EAGRTYRGIAVLGHYLRQRPPSSFLKRIKRTIASDPPRFFNVTDALDIYDALGD 352
 195 QAG-----IA----- 199
 353 RIFGLEGSHAEENESSFQLEMSQIGFSTRHLKDGILFGWAGVADMGGSVLMLEGRHLP 412
 200 -----GFTTEEL---VVMGAPGSTYMACTIKVNLTDNTY-- 231
 413 RMALEDEPPPALQNTAAYIGYSVSSMLRGGRRL-FLSGAPRFRHKGVIYAFQI-KKQGA 470
 232 -LKLNDDEV--IMNRRTYLVGAVTAGHFSHPSTIDVVGAPQDKIGKVIYFRADRRSGT 288
 471 VRVAOSLQGEQIGSYFGSELCPDLDROGTTDVLVLAAPMELGPKNKEGRYVYVVLVQ 530
 289 LKIKTQASGKMGKSGFSSGLCAVDLNGDGLD-LLVGAPMF--SIRREGQVTVYINRGN 345
 531 SLTLTGTLQPPPPDARFGAFMGALPDLNDGPFADVAVGALPDGHOAGATLVYGTSG 590
 346 GALEEQALVLTGGAANAHGESIASLDDLDNDGDFDVAIGAKKEDDFAGAVVIYHGDAG 405
 591 VRPHPAQRIAAASMPHALSYFGRSYDGRLLDGDLDVVAVGA--QGAAILLSRPVYHL 648

406 IVPQYSKMLSGQKINPVILRMFGQISGIDMDNGYDPDVTVGAFMSDSVLLRARPVITV 465
 649 TPSLEVTQPAISVQQRDRRGQAVCTALACF-----QVTSRTP 689
 466 DVSI-FLPGSINITAPQCHDQOQPVNCLINTTCSFHKAVPEELGLNYVLADYAKK 524
 690 GRMDHQFYMRFTASLDEWTAGARAAPDQSGGRLSPRLRLSVGNVTCQOLHFHV-LDTSD 748
 525 GQMPREVYFVLLGETMGQVT-----EKLQTYMBETCHYVAHVYRQVD 568
 749 YLRPALVTAL-----DNTTKP-GRVL--NEGSPISIQKLVPSKQCPDNECVT 797
 569 VISPIVFEAPASLEBHVTEBERELPPLTPVLRKKKGQKIAQKQVTFEERNCRS-DCAA 627
 798 DLVQVNMNDIG-SKAPFVVRGGRKVLVSTLENKRNAYNLSLTIFFRNILASLT 856
 628 DLQIQKLLSMBEKTYLALGAVKINSLNISINLQDDYADVANSFVNSRELFFINMW 687
 857 PQRSPPIKECAPSAHARLCSVGHVVFQTAGKVTFLLEFFSGSSLSLOVFGKL-TASS 915
 688 QKEE--MGISCELLESDPLKCSGVFPFMRSKYEFVIFPTSHLSCBEEVLSFIVTAQS 745
 916 DSLERNGLQNTIQTSAIYQYE-----PHLFSSEST-----LHRYEVRYG 958
 746 GNTRESLSLHNTLVLMVPLMHEVDTSITGMSPTSFYVGSVDANFIQDDBECH--- 802
 959 TLVPGGPEFTTLRVONLGCYVSGLIISALP-AVAGSGNYFLASQVIT--NNASCI 1015
 803 FQPI-----NITQVNYTGPSTLPGSSVSISFNRRLSSGAEHFNQEMVVGQEKNGS 856
 1016 VQNTPEPPRPVHBELOHT-----NRLNGSNTOCQVVRCHLQOLAGTEVSV 1063
 857 FQKPTPCIIQOQENIHHTIFAFFTKSGRAVLDCERKGISCLTAHGNFSAIAKEESRTI 916
 1064 GLILVANE-----FFRAKFK--SLTVSTIFELGTEGSLQLTESRMS 1109
 917 DIYMLNTEILIKDSSSVIQCMSRAKVKVDALVEIAHGNPREVTV--VFEA----- 968
 1110 LLEVQTRPILISWILIGSVLGLLALLVFCMLKGF--AHKKIPEEKKEKLE 1166
 969 LHLNFBRYVVG-WIATISLLVGLIFLLAVLMLKGFRRRYKEIIEAKRKENE 1025

RESULT 13
 JC7294
 alpha9 integrin - sea urchin (Strongylocentrotus purpuratus)
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
 C:Accession: JC7294
 R:Suban, J.M.; Just, M.L.; Lennarz, W.J.
 Biochem. Biophys. Res. Commun. 272, 929-935, 2000
 A:Title: Cloning and characterization of alpha9 integrin in embryos of the sea urchin St
 A:Reference number: JC7294
 A:Contents: Embryo
 A:Accession: JC7294
 A:Molecule type: mRNA
 A:Residues: 1-1054 <SUS>
 A:Cross-references: GB:AD55724
 C:Genetics:
 A:Gene: sualpha9
 C:Superfamily: Integrin alpha-2b chain
 C:Keywords: calcium binding; embryo; glycolysis; glycoprotein; heterodimer; transmembran

Query Match 12.0%; Score 732.5; DB 2; Length 1054;
 Best Local Similarity 23.4%; Pred. No. 5.3e-46;
 Matches 294; Conservative 188; Mismatches 432; Indels 345; Gaps 47;

10 LFLPLVFLTGL-----CSFNLDEHNRPLFPGRPEAFEGSVLQHVGGGQRMVLG---AP 61
 4 FLLSFTCTVLVLDSTYAFNFDLRAVYKFDGQSLGFSVAOHRDQNTDWTVLGABAP 63
 62 WDGSGRGDRGVYRCVVG--AHNAPCAK---GHLDGYQLGNSSHPAVNMHLGMSLETDG 117

```

Db      64 TTQGVNNGAVKCVPTPLSGSGCEQVFPDITGTEVLDSK-----NQMFATLASGP 119
Qy      118 DGFMAAP--LMSRACGSSVFSSGICARVDASFQSGSLAPRACRPTMYDVIVLDS 175
Db      120 DGIILACAPRLVWLQT-----STISPT----- 141
Qy      176 NSIYPMSEVQTFRLRIYVKLFIIDPEQIQVGLVQGESPVHWEMLGDPRTKEEVRAAKUL 235
Db      142 ----- 141
Qy      236 SRREGRTYQAQIIVACTEGFSQSHGGRPEARLLVVTGESHDEELPALKACENG 295
Db      142 -----DKEREPT----- 148
Qy      296 RYTRGIVAVLGHYLRQRDPSSFLREIRTIASDPDERFFPNVTDEALDIYDALGRIF 355
Db      149 -----GTCFVGH-----SDFTNFVNSPCOSTDRD-----LF 175
Qy      356 GLEGSHAENESSFGLMSQIGFSTHRLKDG--ILFGMVGAYDGGSVLWLEGHRLFPPR 413
Db      176 GPD-----KITHCQAGFSAQIPSDNSTIYMAKPSY-----YLOG--QIFAGS 216
Qy      414 MA-LED-----EPPALQNHAAVLYGVSSSMLLRG--GRRLFLSGARFRFH--RQKVIAPQL 465
Db      217 LSTLSDVSNTPEDQAVAFDN--SYRGYSALAGDFNGDGLDYYVGTFRGESLRGLVAIFD- 273
Qy      466 KQDGAVRVQSLQGEQIGSYFSELCPLDTRDGTVDVLLVAPMFL-----GQNKETGR 521
Db      274 --OSLVEITTPVVGQIVSYFSGSVASVDVNGDGLD--LLVGAPMTNREBATEKWEAR 330
Qy      522 VVYVL-----VQOQSLTLQGTLOPEPPODARFGFAMGALPDINODGADVAVAGALE- 574
Db      331 VVYVLQNMASHSGAPQMLTGKTI-----RARFGPITISIGSNDGDNDAIGAPYVG 383
Qy      575 DGHQGLVLYHGTQSGVRPHAPRIAAASMPHA--LSYFGESVDGRLLDGDLDVAVGA 633
Db      384 ENSGGVVYLYHGSABEGLRLTESQVLTLPSELGFSDITTFGSSVGGQDMQNDVPLVGA 443
Qy      634 QG--AAIILSSRPIVLTSLSEVTPQAISVVGDRCK--RRQGEAVCLTALCFQVT--SRFP 689
Db      444 ESADAALIVTRRVLLLEBELTEPIGINDKNTYLPDGTMYTSVANAACFYITGNLIP 503
Qy      690 GRWDHOFYRFTASLDEWTAGAARAFDSGQRLSPRLRLSVGNVTCQDHLHFVLDTP--SD 748
Db      504 ARIGISYTLTVDSI--TSGRALLLEVNBLSQVTKNRMLDYGMKRCDCDLRAVAVNTID 560
Qy      749 YLRPAVLTVPALDNTT-----KRGVLAEGSPTSIQKLVPSKCGPNECVTDLVLOY 803
Db      561 KLTPLAVDQVLTDESILPEILPIINKEAVSQTQVSIQNNC--VNNICIPELGITV 619
Qy      804 NMDIRSRKAPFVVRGGRKVLVSTLENKENVNTSLIIFSRMLHLASLTPKRESPI 863
Db      620 TPNL-----PNIVQAOBELTLVSIINNKGDAFQSTLAVVYPBQLQYRLERRANMF 673
Qy      864 KYECAPASAHARL--CSVGHVPVQFGAKVTFLEEFESCSLSQVFGK-----LTASS 915
Db      674 SYTCTEDSLRMITCTGNPLVG-----KYNLEFGILTSL--QVSGDMONIEFYLVAAS 726
Qy      916 DSELRNGTIOENTAOISAYIOYEP--HLFSESTIHRVYHAPYGLPLVGP----- 964
Db      727 ENNEDBNITLNNELNANTVAIVDATLKLISASYPEIVTYRVPEDNIVPEFPKNASEADI 786
Qy      965 GPEFKTTLRVONLGCVVVSGLIISALLPVAHAGNFELISOVITN--NASC--IVONLTPR 1022
Db      787 GHEVHLHYEVNRTGSSNAAEVTLINRWPEKENDYLYLLGIMDEGVTCQISQOQANP 846
Qy      1023 PGPVPAPEELQHTNRLNGSNTO-----COVVRCHIGQL 1055
Db      847 LGVKL--EASTKEQJNSGTTOVSGRKRKEGEVAAEALQAAPIFCPPECVCLINCTIDBI 903
Qy      1056 AKTEVSVGLRLVINEFFRRAKFKSLTVVSTFELTEGGSV-----LQLTASR 1105
Db      904 KATKSKVIRILGRFMRTEFOKAVSEAVPIYOVTLASTATATVRSIPIYNIPLIMEFTDSTK 963

```

```

Qy      1106 WSESL-LEAVQTRPRLTILSLMILIGSLGLLALVFLCWLKLGFFAKKTIPEEKRE 1163
Db      964 ASTLTABELVPPVSIAMWIIIVSVLSGLIILLIILIGLWKCGFERKKFGEEDKYE 1022

RESULT 14
A41131
lymphocyte-Peyer's patch adhesion molecule alpha 4 precursor - mouse
M/alternate names: Integrin alpha-4
C/species: Mus musculus (house mouse)
C/date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 24-Sep-1999
C/accession: A41131; S16742
R/Neuhaus, H.; Hu, M.C.T.; Hemler, M.E.; Takada, Y.; Holzman, B.; Weissman, I.L.
A/title: Cloning and expression of cDNAs for the alpha subunit of the murine lymphocyte-
A/reference number: A41131; MUID:92064645; PMID:1840602
A/accession: A41131
A/status: preliminary
A/molecule type: mRNA
A/residues: 1-1039 <NEU>
A/cross-references: EMBL:X53176; NID:951484; PIDN:CAA37316.1; PID:951485
C/superfamily: Integrin alpha-4 chain
C/keywords: cytoskeleton; transmembrane protein

Query Match
Best local similarity 22.7%; Pred. No. 5e-43;
Matches 286; Conservative 186; Mismatches 396; Indels 392; Gaps 59;

Qy      23 FNLDEHPRRLPPGPEAFGYSVLQHVGGQRMVLVGP--W--DGPSGDRRGDYRCPV 78
Db      41 YLVDENALLVQGPBGTLFGYSVLHSHGSRKMLVGPATSMLSNANVPALYRCGI 100
Qy      79 -----GGAHAPKAGHLGDIYQUNSHRAVNMHLGMSLLETQD--CGFMAC 124
Db      101 RKNPNQCEQLOSQSGSPGPGKTCLEERD-----NOMLVTLRSQPEGNSIVTC 151
Qy      125 APLMRACGSSVFSFGICARVDASFQSGSLAPTRACRPTMYDVIVLVDGNSIYPMSEV 184
Db      152 GHRW-----KNIF-----YMKSDNKLPT----- 169
Qy      165 QTFRLRVLQKLTIDPEQIQVGLVQGESPVHWEMLSGDRTKEEVRAAKULSRREGRET 244
Db      170 -----GICVMP-----SDLRTE-----LSKR----- 186
Qy      245 TQAQIMVACTEGFSQSHGGRPEARLLVVTGESHDEELPALKACEAGRVTRYGIAY 304
Db      187 -----MACYKDYTRKFGEN-----FASCOAG----- 208
Qy      305 LGHYLRQRDPSSFLREIRTIASDPDER-----FFPNVTDEAALTDIVDALGRIFGLE 358
Db      209 -----ISSFTQULIMGAPGSSYWTGYTVVNIIT-----TNQYKAFVDR----- 248
Qy      359 GSHAENESSFGLMSQIGFSTHRLKDGILFGMVGAYDGGSVLWLEGHRLFPPMALED 418
Db      249 -----QNVKFG----- 255
Qy      419 EPPALQNHAAVLYGVSSSMLLRGRL--FLSGAPRFRHGRKVIAPOLKQDGAVRVQSL 477
Db      256 -----SYLGSVAGHFRSPHTTEVVGAQOHEQIGAYIFSI--DEBELIYVEM 304
Qy      478 QGEQIGSYFSELCPLDTRDGTVDVLLVAPMFLGPON--KETGRVVYVVLVGOQS--IL 533
Db      305 KQKKGISFGASVCAVDADNADGFSQ--LLVGAPM-----OSTIREBGRVVIYINSGGAMV 359
Qy      534 TLQGTLOPEPPODARFGFAMGALPDINODGADVAVAGALEDHQAGALYLYHGTQSGVR 593
Db      360 EMERVLVGSDDKYARAFGESIANLGDIDNDGFEDIALIGAPQDDDLRGAVIYINGRVDGIS 419
Qy      594 HPAQRIIAASMPHALSYGRSVDRGLDLDGDLDVAVGA--QGAAILSSRPVHLTPS 651
Db      420 TYSQRIEQQQSLSLRMGQISIGQIDADNNGYDVAVAGAFQSDSAVLLRLRPVIVVAS 479

```

QY 652 LEVTPQAISVVDRCRRRQGEAVCLTALCFQVTSR-TPGRMDHQFYKFTASLDE---- 706
Db 480 IS-HBESVNRKFDCTENGLPVCMTLTLCFSYKGEVPG---YIVLFYNNLSLVHRKA 534
QY 707 -----WTAGARAAFDGGGRLSPRRRLSVGNVTCQLH--FHVLDTSYLRPVALT 756
Db 535 ESPSRFYFSNGTSDVITGS-----IRVSSSGEKC-RTHQAFMRKDVRDILTFPIHVE 585
QY 757 VTFALDN--TTKPG-----PVINEGSPTS-IOKLVFSDKCPDNECVTLVLOVMN 805
Db 566 ATYHGHVITKRNTEEPPLQPILOQKKEXDVRIMINFAFCAYEN-CSADLVQSAKV 644
QY 806 D-IRSGRAKPFVVRGRRKVLVSTLENRKENAVNTSLIIFSRMLHLASLTPOQESPLK 864
Db 645 GFLKPEYNTKYLAVGSMKTMNLNVSLEFNAGDAYETTLNVQLPTGLYFPIKILDEBKQIN 704
QY 865 VECAPSAHARL-CSVGHVPFQTGA--VTFPLEFEPSCSSLSQVFGKLTAS-SDSLER 920
Db 705 CEVTSSSGIVKLACSLIGIYVDRLSRIDISFLDV-----SSLSRHEDLSIVHASSEN 759
QY 921 NGTL---QENTAQTSAYIOYEPHLLFSSSESTLHRYEHP---YGTLPVPGPE----- 967
Db 760 BEELQOVNRNRYTLTILPLATEVWL-----TVHGL-VNPTSPVYGSSEENE-PETCMAEK 811
QY 968 FTTLRVQNLGCVVVGSLIISALLPAVAHGNYFL-----SLSQVITNNASC----- 1014
Db 812 LNLTFHVINTGISMAPVSVKIMVP-----NSFLPDQDKLFNVLDVQTTGQCFHKHYG 865
QY 1015 -----IYQNLTEPPGPVHPELOHTNR-----LNGSNTQOVVRCHIGOLAKGTE 1060
Db 866 RECTAQQKGIAGTLTD-----IYFSLKTDKRLLYCMKADQHCIDPLCNGKMSGKE 919
QY 1061 VSVGLRLVHNFPRRAKFSKSLTVVST--FELGT-----EEGSLVQLTEASRMSLSLE- 1112
Db 920 AS-----VHIOLEGRPSLTLEMDETSLSLKEIKATAFEPHPKRVKLEINDEVAVHFLDG 973
QY 1113 VVQTP-----ILISLMLIGSVLGGLLALLVLC-LWKLGF--AHKKTPEEKRE 1163
Db 974 LHHQPRKHFHTIIITISLLG-----LIVLLLSICVMWKAGFPKQKSLIQEENRD 1027

RESULT 15

S06046
integrin alpha-4 chain precursor - human
N:Alternate names: CD39d; very late antigen-4 alpha chain; VLA-4 alpha chain
C:Species: Homo sapiens (man)
C>Date: 01-Dec-1989 #sequence revision 01-Dec-1989 #text_change 24-Sep-1999
C:Accession: S06046; A39355; D28018
R:Takada, Y.; Ellices, M.U.; Crouse, C.; Hemler, M.E.
EMBO J. 8, 1361-1368, 1989
A>Title: The primary structure of the alpha(4) subunit of VLA-4: homology to other integ
A:Reference number: S06046; MUID:89356603; PMID:2788572
A:Accession: S06046
A:Molecule type: mRNA
A:Residues: 1-1038 <TA>
A:Cross-references: GB:X16983; EMBL:X15356; NID:G33945; PIDN:CAA34852.1; PID:G33946
R:Rosen, G.D.; Birkenmeier, T.M.; Dean, D.C.
Proc. Natl. Acad. Sci. U.S.A. 86, 4094-4098, 1989
A>Title: Characterization of the alpha-4 integrin gene promoter.
A:Reference number: A39355; MUID:91239513; PMID:2034655
A:Accession: A39355
A:Molecule type: DNA
A:Residues: 1-93 <RNA>
A:Cross-references: GB:M62841
R:Takada, Y.; Strominger, J.L.; Hemler, M.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 3339-3243, 1987
A>Title: The very late antigen family of heterodimers is part of a superfamily of molecu
A:Reference number: A94151; MUID:87204112; PMID:3033641
A:Accession: D28018
A:Molecule type: protein
A:Residues: 40-50, 'E', 52-53 <TA>
C:Genetics: GDB:ITGA4; CD49D

A:Cross-references: GDB:128032; OMIM:192975
A:Map position: 2q31-2q32
C:Superfamily: integrin alpha-4 chain
C:Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; transme
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-1038/Product: integrin VLA-4 alpha-4 chain #status predicted <MAT>

Query Match 11.2%; Score 684; DB 2; Length 1038;
Best Local Similarity 22.9%; Pred. No. 1.4e-42;
Matches 289; Conservative 181; Mismatches 423; Indels 368; Gaps 54;

QY 9 LFLPLVFLTGLCSPPNDENHRLPPRPEAFEGYSVLQHYGGGRMVLVGP--W-DG 64
Db 28 LLLCIGVPTG--RPYNDTESALLYQGPHNTLFGSVVLSHGANKRMVLVGAPTAMVLAN 85
QY 65 PSGRBRGVNRCPVY-----GAINA-PCAKGLGDYOLGNSHPANMHIGMS 111
Db 86 ASVINPQAIYRCRGNKPNQGTCELOLQSPNGEPQKTCLEERD-----NOMLVGT 136
QY 112 LLETGDG-GGFMACAPLWSRACSSVFSSGICARVDASFPQGSILAPTRQCPYMDVVI 170
Db 137 LSRQGENGSIVTCGHR----- 154
QY 171 VLDSGNSIYPMSEVOTFLRLVGLKFLIDPEQIOGLVOYGESPVHMSLGPRTKEVVR 230
Db 155 ---KNIFYINR-----NKLPTGGC-YGVPP-----DLRT----- 180
QY 231 AAKSLRSREGETTAQAIWVACTEGFSQSHGSRPEARALLVVTDSHSGEELPALK 290
Db 181 -----ELSKRIA 187
QY 291 ACEAGRVYRGIAVLGHYLRQRDPSSFLREIRTIASDPDERFFFNVTDEALTDIVDAL 350
Db 188 PCYQDYVKKF----- 198
QY 351 GDRIFGLEGAENESSGLEMSTQIGSTHRLKQGLFGMVGAYDWGSYMLR---CGH 407
Db 199 -----ENFAS-----COAGISSFYTKDLIWAGAGSSYWTGSLFVYVNTTKY 241
QY 408 RLFPFRMALDEPPALONH---AAYLGYSVSMLLGGRRL-FLSGAPRRHRHGKVIAP 463
Db 242 KAFIDK-----ONOVKFGSYLGYSVAGAFHSQHTTEVAVGAPQHEQIGKAYIF 290
QY 464 QLKDDGAVRVAQSLQGEIGSYFSGSELCPDLTDRTDITDVLVLAAPMELGPON--KETGR 521
Db 291 SI-DEKELNIIHEMKKKLGSYFGASVCAVDLNDGSSD-LVGAAPM---QSTIRREGR 344
QY 522 VYVYL-VGQOSLL-TLQGTLOPPPODARFGFAMGALPDLNQGPADYAVGAPLEDHGQ 579
Db 345 VFVYINSGGAVNMMETNLVGSDKYARFGEISIVNLGDIINDGFEDEVAIGAPQEDDLOG 404
QY 580 ALVLYHGQSVRRHPRORILAAASMPHALSYFGGSVQRCRLDDDDLDVDAVGA--GQA 637
Db 405 ALIYINGAGDISSTFSQRIEGLQISLSLWFGGISGQIDADNNGYDAVAVGARSDSA 464
QY 638 ILLSSRPVHHTPLSELEVPOAISVVDRCRRRQGEAVCLTALCFQVTSR-TPGRMDHQF 696
Db 465 VLLKTRPVVLYDALSS-HBESVNRKTFDVCVNGMPSCIDLTLCFSYKGEVPG---YI 519
QY 697 YMRPTASID--EWTAGARAAFDGSG-QRSFRRRLSVGNVTCQLH--FHVLDTSYLR 750
Db 520 VLFYMSLDVNRKAESPFRFYFSNGTSDVITGSIQVSSRANC-RTHQAFMRKDVRDIL 578
QY 751 RPAVLTVTFAL-----DNTTKGPVINESSPSI-QKLVFSDKCPDNECVTL 799
Db 579 TPIDIEAAVYHGHVVISKSTEEPPLQPILOQKKEXDVRIMINFAFCAYEN-CSADL 637
QY 800 VLQVQMD-IRSGRAKPFVVRGRRKVLVSTLENRKENAVNTSLIIFSRMLHLASLTPO 858
Db 638 QVSAKIGFLKHEKKTLYAVGSMKTMNLNVSLEFNAGDAYETTLHVLKPLVGLYFIKIL 697
QY 859 RESPIKVECAPSAHARL-CSVGHVPFQ--TGAKVTFLEFEFSGLLSQVFGKLT-- 912

Db 698 EEKQINCEVTNDNGVVQLDQSIGIYVDHLSRIDISFLLDV-----SSLRAEEDLSITV 752
QY 913 -ASSDSELRNGTLQENTAGTSAVIOYEPHLLFSSSESTLHREYHP-----YGTLPVGPGE 967
Db 753 HATCENEEMDNKHSRVTVAIPKYEVL-----TVHGF-VNPTSFIYGSNDENE-PE 804
QY 968 -----FKTLRVONLGCYVVSGLISALLP-AVAHGNVFLSLSCVITNNASCIQONLT 1020
Db 805 TCWVEKMNLTPIHINTGNSMAPVSVIWPNSFSPQTDKLFNILDVQTTGEC----- 858
QY 1021 EPPGPPVHPPELOHTNRLNGSNTQCQVR-----CHLG 1053
Db 859 -----HFENYQRYCALCQOKSAMQTLKGIYVFLSKTDKRLIYCIKADPHCINFLCNFG 911
QY 1054 QLAKGTEVSVGLRLVHNEFFRRAKFKSLTVST--FEL--GTEEGS--VLQLTEASRW 1106
Db 912 KMEGSGEAS-----VHIQEGRPSTLEMDETSALKFEIRATGFPPEPNPVIENKENV 965
QY 1107 SESLLEVV--QTRPILISIMILIGSVLGLLIALLVFCIMKLGFF--AHKKIPEBEKRE 1162
Db 966 AHVLEGLHQRPKRYFTVIISSELLGLIVLLISYMMKAGFFKQYKSIQJENNR 1025
QY 1163 E 1163
Db 1026 D 1026

Search completed: July 16, 2003, 07:56:45
Job time : 67.9093 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

COM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 119.819 Seconds
(without alignments)
2006.842 Million cell updates/sec

Title: US-09-647-544-2
Perfect score: 6106
Sequence: 1 MELPVTHTLPLVFLTGLC.....GFPAHKKIPBEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853.5	30.4	1171	13 042094	042094 gallus gall
2	1487.5	24.4	823	4 08MY18	08MY18 homo sapien
3	1136	18.6	1161	11 09QY17	09QY17 rattus norv
4	1097	18.0	1161	11 09MT04	09MT04 mus musculu
5	1089.5	17.8	1160	11 09R200	09R200 mus musculu
6	1069	17.5	1169	11 09QX04	09QX04 mus musculu
7	1000.5	16.4	1151	11 09J130	09J130 rattus norv
8	997	16.3	780	13 006271	006271 xenopus lae
9	984	16.1	1086	4 096H01	096H01 homo sapien
10	980.5	16.1	1196	13 098T01	098T01 cyprinus ca
11	953	15.6	1167	11 088340	088340 rattus norv
12	947.5	15.5	1187	13 098T00	098T00 cyprinus ca
13	946	15.5	1167	11 088341	088341 rattus norv
14	883.5	14.5	920	6 028984	028984 sus scrofa
15	883	14.5	895	11 09WU08	09WU08 mus sp. itg
16	797	13.1	1332	5 09BP08	09BP08 halocynthia

17	778	12.7	1036	11 091YD5	091YD5 mus musculu
18	732.5	12.0	1054	5 09UG61	09UG61 strongyloce
19	725	11.9	1033	6 09BG31	09BG31 bos taurus
20	691	11.3	1032	11 061989	061989 mus musculu
21	681	11.2	1034	13 098RT7	098RT7 gallus gall
22	661	10.8	1041	5 076378	076378 lyechinus
23	655	10.7	1041	5 09UB90	09UB90 lyechinus
24	623.5	10.2	1033	13 042598	042598 xenopus lae
25	583	9.5	1016	13 091779	091779 xenopus lae
26	576.5	9.4	974	11 0924W2	0924W2 rattus norv
27	570	9.3	1045	13 091292	091292 pleurodeles
28	562.5	9.2	1049	4 096H05	096H05 homo sapien
29	560.5	9.2	1047	6 09MZD6	09MZD6 bos taurus
30	551.5	9.0	1007	6 09GK48	09GK48 bos taurus
31	551	9.0	833	5 09BP07	09BP07 halocynthia
32	536	8.8	1034	6 09TUN4	09TUN4 cycloclagus
33	524.5	8.6	1036	6 09TUN6	09TUN6 sus scrofa
34	516.5	8.5	1012	11 070304	070304 mus musculu
35	515.5	8.4	1049	5 08SY51	08SY51 drosophila
36	511	8.4	1087	13 091778	091778 xenopus lae
37	469	7.7	1036	6 09TUN8	09TUN8 canis faml
38	461	7.5	1036	6 09TUN4	09TUN4 canis faml
39	450	7.4	1021	5 09GSP4	09GSP4 podocoryne
40	438	7.2	1086	5 018428	018428 geodia cydo
41	394.5	6.5	191	6 029124	029124 sus scrofa
42	374	6.1	764	6 028762	028762 papio cynoc
43	353	5.8	622	13 091570	091570 xenopus lae
44	305	5.0	673	5 09SYW0	09SYW0 polyandroca
45	251	4.1	1472	13 090Z00	090Z00 gallus gall

ALIGNMENTS

RESULT 1

ID	042094	PRELIMINARY;	PRT;	1171 AA.
AC	042094;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	ALP1AI integrin.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISUE=GIZZARD;			
RX	MEDLINE=97476270; PubMed=9334246;			
RA	Obata H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,			
RA	Sobue K.;			
RT	"Smooth muscle cell phenotype-dependent transcriptional regulation of			
RT	the alpha integrin gene."			
RL	J. Biol. Chem. 272:26643-26651 (1997).			
DR	EMBL; AB000470; BAA33160.1; -			
DR	EMBL; AB000471; BAA33161.1; -			
DR	HSSP; R17301; IAOX.			
DR	InterPro; IPR000413; Integrin_alpha.			
DR	InterPro; IPR002035; VWF_A.			
DR	Pfam; PF01839; FG-GAP; 5-			
DR	Pfam; PF00357; Integrin_A; 1.			
DR	Pfam; PF00092; vwa; 1.			
DR	PRINTS; PR00453; VWFADOMAIN.			
DR	SMART; SM00191; Int_alpha; 4.			
DR	SMART; SM00327; VWA; 1.			
DR	PROSITE; PS00242; INTEGRIN_ALPHA; 1.			
DR	PROSITE; PS50234; VWFA; 1.			
KW	Integrin.			
SQ	SEQUENCE 1171 AA; 130228 MW; B505AD65F09736E CRC64;			
Query Match	30.4%;	Score 1853.5;	DB 13;	Length 1171;

Best Local Similarity 34.7%; Pred. No. 2,5e-131;
Matches 416; Conservative 226; Mismatches 464; Indels 83; Gaps 21;

```
QY 13 LVFLTLCSPNIDENHPRLPFPPRAEFGYSLVGHVGGQRMVLGAWMDPSGRRD 72
Db 8 LVALPHLASVNDVKNAMTFSGPLEDMFGYVQOQENEGKVVLLGSLVLPQPEKRTD 67
QY 73 VYRCRPGAHNAPCAKGLGD-YOLGNSSHPAVNMHLSLETDDGGFMACAPLMSRA 131
Db 68 VYKCPGRDSQSPCITLNPDATSVPNWVEKNMMLGTTIV-TNKXGFLACGPIYAK 126
QY 132 CESSVSSGICARVDASFOPOGSLAPTAQRCPYMDVIVLDGNSIYPMSEVOFRLRL 191
Db 127 CRLHYTTGCVSVSSTFETVKAVAPSOECKTOLDIVLDGNSIYPMSEVTAFINSL 186
QY 192 VKGLTFDEPOIOVGVGVESPVHEWSLGDFTKEVAVNAAKNLSEBREFKTAQAINV 251
Db 187 LKNMDIGPOOTOVGVQTVVHEFYLTSTTEEMDAALRIRQGGTQMTALGIDT 246
QY 252 ACTEGSGHGGPRPEAARLLVVTVDGSHDEELPALKACEAGRTYRGIAVLGHYLR 311
Db 247 ABEATEAHGARQGVQKMWIVTDESIDNRLQVICKEDENIQRPAIALISYSG 306
QY 312 QRPSSFLREIRTIASDPERFFNVTDPAALTDIVDALGDRIFGLESGHAENESSFGL 371
Db 307 NSTEKFEVEIKSIASKPTEKEHFNVSDLAITVIVEALGERIFALEATTDOOASFME 366
QY 372 MSQIGSTRRLKDLGIFGMVAVDMGGSVLMLEGGHRLPPRMALDEFPALONHAAYL 431
Db 367 MSQAGSAHVSQDWWLGAAGVADMGVVMVWDSDISIPNDTFDRHSEKIEPIAAL 426
QY 432 GYSVSMILRGGRRLFLSGAPRFHHRGKVIAPOLKKGAVRAVQSLQGEQISYFSGELC 491
Db 427 GYTVASALPRPG-VLYIAGOPRNHGVYIYKM-EGREVKVLQRLKGGQISYFSGVIT 484
QY 492 PLDTDRDGTTLVLLAAMPFLGPONKETERVYIYLVG-----QOSLTLQGT----- 538
Db 485 TIDNRDSFTDLLVGAFTYMGTEKEEGKVYVVALNKTKEFQMSLEPIKOTCCSPLHG 544
QY 539 -----LQEPPODARCFAMGALPDLDNQGFADVAVAPALPDEHQGLYLYHGTQSVRPH 594
Db 545 DTCKVLKNEPCGARFETAIAAKVLDLNDGNDIVIGSPLEDDHGRVVIYHGHGNTISK 604
QY 595 PAQRIAAAMPALSYFGSRVSDGRDLDDDDLDVDAVGAQAAILSSRPVHLTPSLEV 654
Db 605 YQRIASGGDEKVFEGSGVHGENMDLNDGLDIVIGLGAALFWSDDVAEVANSQF 664
QY 665 TPQAISVVDRCRRRQGEAVCLTALCFQVTSRTPEGRMD-HQFYMRFTASLDWTAGARA 713
Db 666 MKXSINIOQONCOINRKTCINATICFK--TRLKSKEDIPESSLDQWITLDAORQISRS 722
QY 714 ARDGGQRLSPRLRLSVGNVTCQOLHFNVLDTSDVLRVALTYVPALDNTTKP--GPLY 771
Db 723 LTFETHERMOK--NTTIKGSBCIKHFPMLDKPDQDSVKVLEF--NFSDPESGPVL 777
QY 772 NEGSPTSIOKLVPFSKCPDNECVTDLVLOVNMDRSGRKA PFVVRGRRKVLSTTLE 831
Db 778 DSNLNSISEYIPTFTDCGAKKNCISDLALNVKASLAGDSSSPFYKSNNDPFTQLSLK 837
QY 832 NKENAVNTSLSIIFSRNHLASLP-QRESPIKVECAAPSAHARLCVGHFVFTGAKV 890
Db 838 NKKDSAYNTRALVQVSPNIIIFAGIEDIQKDS-----CESNNHITCKGVFPFLKEEBI 890
QY 891 TLLLEPEPSSLLSGVFGKLTASSDSELRNGTLQNTQTSAYIOYERHLLFSSSSTIH 950
Db 891 SKRISFOFASVLLLENATDVYATSDSEBPETLRNRRGQVITI PVYEEGLLFVSVFEKH 950
QY 951 RYEVHPYGLPVG-----PGPEFKTTLVONLGVVSGLLIISALLPVAHAGN---YF 1001
Db 951 HYIIAANETIPTAINTTEOIGDEVTLHYIEKGEHPRMNLTLQILYPRVTAKTYLXI 1010
QY 1002 LSLGVITNNASC-----IVONLTPEPCPVHBEIOLHTRNLNGSNTQ 1044
```

```
Db 1011 TLLSH--SQMAICKSSYPVDHLKIGSGKSVYLPKIKEPTKDTI-----MECDTFS 1058
QY 1045 COVVRCHLGQAKGNEVSGLLRLVHNEFPRAKRSITVSTPELGEBSVLQLTES 1104
Db 1059 CASINCALAP-SDISQVNSL--RWKPTTIKASHSLTLVVKALLRENSSLIRNDHQ 1115
QY 1105 RMSSESLLEVQTRPILSLIMLIGSVLGLLALLVLCMLKGFPAHKKIPEEKREB 1163
Db 1116 KLEWTIKISKEPPEPSVPLWVILPISIFAGLIILALILFALWKAGF---KRLPKKMEK 1171
```

RESULT 2

```
Q8WY18
ID Q8WY18 PRELIMINARY; PRT; 823 AA.
AC Q8WY18;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE MSTR018.
GN MSTR018.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=AORTA;
RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
RA Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
RA Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
RA Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF111799; AAL39001.1;
DR InterPro; IPR00413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 4.
DR PRINTS; PR0185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 4.
SQ SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;
```

Query Match 24.4%; Score 1487.5; DB 4; Length 823;

Best Local Similarity 39.6%; Pred. No. 8,2e-104;

Matches 327; Conservative 151; Mismatches 308; Indels 39; Gaps 13;

```
QY 372 MSQIGSTRRLKDLGIFGMVAVDMGGSVLMLEGGHRLPPRMALDEFPALONHAAYL 431
Db 1 MSQIGFSHVVEDGVLGAAGVADMGVVMVWDSDISIPNDTFDRHSEKIEPIAAL 60
QY 432 GYSVSMILRGGRRLFLSGAPRFHHRGKVIAPOLKKGAVRAVQSLQGEQISYFSGELC 491
Db 61 GYTVASVSSRQGRVYVAGAPRFHHTGKVIILFTMHNNRSLTIHQMRGQOIGSYFSGEIT 120
QY 492 PLDTDRDGTTLVLLAAMPFLGPONKETERVYIYLVGQOSLTLQGTLOPEPP-ODARFG 550
Db 121 SVDIDGDDVTTLVLLGAMFYRN-BGEREGKAYVEL-RQNPFYVINGTLKDSHSYNARFG 178
QY 551 FAMGALPDLDNQGFADVAVGAPLEDDHQAALYLYHGTQSVRPHPAQRIAAAMPALSY 610
Db 179 SSIASVRLNDSDVNDVVGAPLEDDHNAAGALYIHRFGFSGSLTKPKQITASELATGQY 238
QY 611 FGRSYDGRLLDGDLDVDAVGAQAAILSSRPVHLTPSLEVTPOAISVVDRCRRRG 670
Db 239 FGCSIHQGLDNEGLDILAVGALGNVILMSRPVQINASLHEPESKINIFHNDCKRSG 298
QY 671 QEAVLTALALCFQVTSRTPEGRMDHQFYMRFTASLDWTAGARAADFSGQRLSPRLSL 730
Db 299 RDACTLAFLCFPTPIFLAPHPQTTGVGIRVATYMDERRRYTPRALHDESGDFTNNAVL 358
QY 731 VGNVTCQOLHFNVLDTSDYLRVALTYVPALDNTTKPQPVINEGSPITSIOKLVPFSKDCG 790
Db 359 SGQELCEHINHVLDITAYVAVKVPFVSYSLEDDH-GPMLDDGKPTTLRLSVSPFWNCN 417
QY 791 PDNECVTDLVLOVNMNDI-----RGSRK-----APVVRGRRKVLVST 828
```

Db 418 EDEHCVDPDLVDARSDLPAMFYCQRYLAKKPAQDCSAVTLSDTTVFIIESTTRQAVNEA 477
Qy 829 TLENKENAVNTSLSIIFSRNLHLASLTPORESPIKVECAPS--AHARLCSVGHVPQT 886
Db 478 TLENNGENAVNTSVLINSQANQFASLIQKEDSDSIECVNERLQKVCVNSVFPFRA 537
Qy 887 GAKVTFLFEFSCSLSQVFGKLTASDSIERNGTLOENTAQTSAYIQYEPHLLFSS 946
Db 538 KAKVAFRLDFEFESKSIFFLHLEIEIAGSDSNERDSTKEDNVAPIRFHLYEADVLFTRS 597
Qy 947 STLHRYEYHVPGLP--VGPGEFPTTLRVONLGCYVSGLLISALPLPAVANGVYFSL 1004
Db 558 SLSHYEYKNSLSERYDIDGPPFCIFRIQNLGLPFIHGMKKITPILATSGNRLYL 657
Qy 1005 SQVITN--NASC-IVQNTLPEPGPPVHPELOHTNRNGSNTOCVVRCHLQOLAKTEV 1061
Db 658 RDLFLDEANTSCINIGNSTEYRPTVE--BDLARAQNLHNSNDVVSINONI--RLVPMQEI 715
Qy 1062 SVGLRLVHNEFFRAKFKSLTVSTFELGTEGSLQLTASRMSSEILEVYQTRPLII 1121
Db 716 NFHLGNMLRSLKALKYKMKIMVMAALORQFHSPIFREEDPERQIVFEISKQEDWQV 775
Qy 1122 SLWILIGSVLGLLTLALVFLCMLKGFPAHKKIIEBEKEEKE 1166
Db 776 PLWITVSTGLGLLTLALVIALMKLGFRSAR---RRREPGLD 816

RESULT 3
Q090Y7 PRELIMINARY; PRT, 1161 AA.
AC Q090Y7
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Alpha D integrin.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Gallatin W.M.;
RT "Cloning of rat alpha D, a novel beta 2 integrin."
RL Submitted (Aug-1997) to the EMBL/Genbank/DDbj databases.
DR EMBL; AF021334; AAF21241.1; --
DR HSSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PRO1185; INTEGRINA.
DR PRINTS; PRO0453; VWFADOMAIN.
DR SMART; SM00193; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Integrin.
KW SEQUENCE 1161 AA; 126600 MW; 2258491A984A705E CRC64;

Query Match 18.6%; Score 1136; DB 11; Length 1161;
Best Local Similarity 30.1%; Pred. No. 7e-77;
Matches 363; Conservative 208; Mismatches 486; Indels 148; Gaps 46;

Qy 20 CSPFLDEHRRPLFGPPFAEFYSYLTQHVGGQRMVLGAPWDGSPGDRGDVYRCVPG 79
Db 17 CHGSNLDVBEPIV--REDASFGQTVQF--GSSR-LTVGAPLEAVAVNQTRLDLC--- 69
Qy 80 GAHNAPCAKGHLGDYQLGNSHPAYNMHIGMSLLETGDDGGMACAPLMSRACGSSVPS 139
Db 70 ---AP-ATGMCQPIVL--RSPLEAVNMSLGLSLVTAIINNAQLLAGGPTAQRACVKNMYAK 123

Qy 140 GICARVDASFOPOGSLAFTACRPT--MDVVIIVLDGNSI--YPMSEVOTFLRLVGLKF 196
Db 124 GSCILLGSLQFIQAVPASMPECPRQEMDIAFLIDGSSINSIQRPQAKDQKDFVXLMGE-F 182
Qy 197 IDEBQIQVLYQGESPPHMSLDFTKEEVRPAKSLRREGRETTAQAIVACTEG 256
Db 183 ASTSTL--FSLMQYSNILKHTFTEFKNIIDPOSIVDPIDVLOGL--TYTAAGIRTVME 240
Qy 257 FSOHSGRPEARLLVTVTDGESH--DGEELPALKACGAVTVYGIIVLHYLRORDP 315
Db 241 FHSKNGSKSKAKLILVTTDQKTRDPLEYSDVIPADKAGIIRALVGCVF---QEP 296
Qy 316 SSFLREIRTIASDDPERFFNVTDAAITDIDALGRIFGLEBSHAENESSFGLMSQI 375
Db 297 TA-LKEMLTISAPRQDHFVKGNPALRSIQROQEKIFALIEGQSSSSSFGHEMSQE 355
Qy 376 GFSTRHLKDGLIFGVGAYVWGSSVLMEGHRLFP-----RMALDEFPPALONHA 428
Db 356 GFSSALTSQDGVLAGVGSFSW-----SGAFLYPPNTRPTPTINMSQEN-----VDMRD 403
Qy 429 AYLGSVSMILRGRRLLFGAPFRHKGVIAPOLKKDGAIVAOSLOEOIGSYRGS 488
Db 404 SYLGISTAVARWKGHSLL--GAPRHQRTGKVIF--TQEARHMRPKSEVRGTQIGSYFGA 461
Qy 489 ELCPDLTDGDTVLVLAAPMFLGPQNKETGRVYVYLV--GQGSLLTLQGTLOEPPDPD- 546
Db 462 SLCSVDVDROGSTDLVLIGAPHY--BOTRGQVSVFPVPRVGRWQCEALTLHGQGHFW 519
Qy 547 ARFGFAMGALPDNLQDGFADYAVGAPLEDHOGALYLYHG--TQSGVRPHPAQRIAAAMP 605
Db 520 GRFGVALTLVDVNGDNLADVAIGAPGEEBRCGAVYIFHGASRLIEIMPSFGQVTSQLS 579
Qy 606 HALYFGSVGRDLDDDDLVYAVAGAGAAIILSSRPYHILRPSLEVTQALSVQRD 665
Db 580 LRLQYFGSLSGGDLTODGLVDLAVAGQGVLLRLSLPLKVLUSTIFAPMEVAKAVYQ 639
Qy 666 CRRR-----GOBAVCTIALCFQVTSRTGRMDH--QFYVRFASLDEMTAGAAAPDG 717
Db 640 CWERTPYVLEAGENTVCLT-----YHKSPLDIGNNGSVRYDALDPGLISRAIFDE 693
Qy 718 SGQRSPRRRLSVGNVTCQQLHFRVLD--TSDYLRPVALTVPAL--DNTYK--GPVLYN 772
Db 694 TKNCTLTNGKTKTLGID--HCEYVKLLLPDCEVDAVSPITLRNLSVLDSDASPRNLHPVLA 752
Qy 773 EGSPTSIQKVPFSKDCQDNCEYTDVLQVNMNIRGSRKAPFVVRGGRKRVLSTLEN 832
Db 753 VGSQDHTTASLPFEKNCKQELLCEGDLGISFNFS-----GLQVLVVGSGPELTVTVYWN 807
Qy 833 RKENAVNTSLSIIFSRNLHLASLTPORES---PIKVEG--AAPSAAHL---CSVGHVPF 884
Db 808 EGEDSYGLVVFYYPAGISYRKYVTGTQOPHYPLRLACEAPPAQOEDLRSSCSINHFIF 867
Qy 885 QTKAKVTFLFEFSCSLSQVFGKLTASDSIERNGTLOENTA--QTSAYIYV----- 937
Db 868 REGAKTTMTMIFDVSYKAFIGDRL--LLRAKASSENKPKDINKAFOLELVKXTVYTLI 925
Qy 938 -----EPHLLFSS-----ESTLHRYEYHVPGLPVGPGGEFTTLRLVONLGCYVSG 985
Db 926 SRODSTNHVNFSSSHGGRQEAHRYVNVLSPD-----KLAAYRVNFWPVLNLGV 977
Qy 986 II-SALLPVAVHAGNYFLSLSQVITNNAACIVONLTPPEPGPVHPELOHTNRNGSNTO 1044
Db 978 AVMDVTLSSPQO-----VSCVSO--MKRPQNPDLTQIQRSVLDGSLAD 1021
Qy 1045 QCVVRCHLQOLAKTEVSVGLRLVHNEFFRAKFKSLTVVSTFELGTEGSLVQL--TE 1102
Db 1022 CLHFRCDIPSDIDDELDFIRGNLSFGWASQTLQEKVLVLSBAEITFDTSVYQLPQOE 1081
Qy 1103 A---SRMSSEILEVYQTRPLIISMLIGSVLGLLTLALVFLCMLKGFPAHKKIIEE 1159
Db 1082 AFLRAQVETTLIEEYVEPI---FLVAGSSVGLLTLALVTLVLYKLGFP----- 1128

QY 1160 KRECK 1164
DB 1129 KROYK 1133

RESULT 4
Q9WTV4 PRELIMINARY; PRT; 1161 AA.
AC Q9WTV4;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Integrin alpha L.
GN ITGAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2J; TISSUE=SPLEEN;
RA Ma R.Z., Teuscher C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065902; AAD25885.1; -.
DR HSSP; P20701; 1LFA.
DR MGD; MGI:96606; Itgal.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP_5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS00234; VWF_A; 1.
DR PROSITE; PS00234; VWF_A; 1.
SQ SEQUENCE 1161 AA; 128240 MW; 86B102F7B209E431 CRC64;

Query Match 18.0%; Score 1097; DB 11; Length 1161;
Best Local Similarity 27.9%; Pred. No. 6.4e-74;
Matches 358; Conservative 187; Mismatches 443; Indels 294; Gaps 49;

QY 12 PLVFLTGL-----CSPFNDEHHPRLFGPPEAFEGSYLVQHGQGMVLGAPWDGDS 66
DB 8 PRLLILGLQFAKAWSYNDTRPTQSLAQAGHFGYQVLIQIDG-----VVVGARPE---- 60
QY 67 GDRRCGVYRCVGGANAPCAKGLGDYQIGNSHPAVNMHIGMSLLETDGDEGEMACAP 126
DB 61 GNTGTGLYHCRSTSEFCQPV-----LHGSNH--TSKYLGMTLATDAKSLACDP 110
QY 127 LMSRAGSSVFSFGICARVDASFQPGSLAFTACRCPT-----MDVVIYDGSNST- 178
DB 111 GLSRTDQNYVYSGLC-----YLPQSLGPMQLQNPAYOECMKGVLDVFLFDGSGQSD 165
QY 179 -YPMSEVQFLRLVGLKFLIDPEQIOVGLVOYGESFVHEMSLDFPTEKEVVAANKLNR 237
DB 166 RKDFEKILFEMKDVWKML--SNYSQFAVQFSTDCRTEFTFDY-----YKQNNKPNV 217
QY 238 RRGRE-----TKTAQAINVACTEGFSQSHGRPREARLLVVTTDGSHDGEELPALA 291
DB 218 LIGSVQPMFLNTFPRAINVVAHVAFKESGARPDATKVLVITDGEADKGNISA---- 273
QY 292 CEAGRTYRGIVLGHYLRORPSSFLREIRTAIDPBERPFENVTDALTDIYDALG 351
DB 274 --AHDITRYITIGKPFVSVQK-----KTLHIFASEPVEEFKILDTTEKLDLFTDQ 326
QY 352 DRIFGLEGSHAENESSFGLEMSQIGFSTRHLKDGILFGWGAVDWGSVYLMLEGHRLRP 411
DB 327 RRIYALEGINRQDLTSFNNELSSGSIADLSKGNHVVAGAKDMAGFLDLR----- 379
QY 412 PRMALED-----EPFALONHAAVYLGVSVMILRGGRLLFSGAPRFRRHGVYAF 463

DB 380 -----EDLGATFVGQEBLTSIDVRGVLGYTVAMWNTSSSRPLAAGAPRYOHVQVILF 434
QY 464 QL-KKDGAVRVAQSLQGEQIGSYFSGELCPIDPTDRDGTDDVLLVAAPMFLGPONKETGRV 522
DB 435 QAPBAGGRNMQOKIEGTQIGSYFSGELCSVDLDDGABELLIPAPLFFGEQ--RGGRV 492
QY 523 VYVIVGQSSLTLOGLQPEPPQD--ARFGFANGALPDLDNODGFADVAVAPLLEDHGQAL 581
DB 493 FTY-ORQSLPFEMVSELQDPCYPLGRFGAIALTALTDINGDRLTVAVAPLFE--QGA 549
QY 582 YLYHGTQGVRRHPQRITAASMPHALSYFGRSVUGRLDDGDDVDVAVAGQAAILLS 641
DB 550 YIFNKGPGGLSPQSPQRIGQAVFPFGIRWFGSHIGVXDLDGDRADVAVGEGRVVLS 609
QY 642 SPTIHLTPSLSEVTQQAISVVOQDC-----RRRGQAVCLTALCFQVTSRTPGRDHOFY 697
DB 610 SFPVVDVYTELFSFSEELPHEVEGYSAREQKIGVILKA--CRIRPLTP-----QFQ 662
QY 698 MR-----FTASLDEMTAGARAAPDQSGQRLSPRRLSVGNV-----TEQLHFNH-L 744
DB 663 GRLLANLSTYLQLDGHRMRSGLPDQSGHELS-----GNTSITPKKSCIDPFHPPI 714
QY 745 DTSVLRPVALTVTALDNTTKPGVVLNBSFT-----SIOKLVPEKSD 788
DB 715 CIQDILSPINVSINPSL-----LEBEGTPRDQGRAMQILRPSIHTVKEIPFEKN 766
QY 789 CGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKVLVSTLLENKENAVYTSLSIFSR 848
DB 767 CGEDKKCEANLTL-----SPARSGPLRLMS--ASLAVEMVLNSGDEAYWRLLDPR 820
QY 849 NLIHLASLTP-QRESPIKVECAAPSANARL-----CSVGHVPQTAGAVTFLEFEFSCS 901
DB 821 GLSPFKVEMLOPHSHMPVSCCELTEGSSLLTKTLKCNVSSPFIKAGQVVS----- 870
QY 902 SILSQVFGKLTASS--DSLENGTL-----QENTAOISAIYQVEPHLFFSESSTL 949
DB 871 -LQVMEFTLLNSSWEDFVELNGTVHCENENSSLOEDNSAATHPLVLPVNLITKEQEN- 927
QY 950 HREYHPYGLTPVCGPGEFFKTLRLVQNLGCYVVSGLIISALLPAVAGNGYFLSLQYIT 1009
DB 928 -----STLYISTPKGPKTQOVQHV-----YQVRQBPAY 957
QY 1010 NNASCIQNLTEPPGP-----PV-----HPEELQHTNR-----LNGSNTQC 1045
DB 958 DHNMPTLEALVGVPMHSEDPITYTWSVQTDPLVTCHSEDLKRPSSBAEQCLPVGQRC 1017
QY 1046 QVVRCHQLQAKGTEVSGLRLVHNEFFRAKFKSLTVSTFELGT--EGSVQLTREA 1103
DB 1018 PIV-----FRREIL--IQVTGIVELSKETIKASSTLSGSS 1050
QY 1104 -----SRMSSES-----LLEVQTRPILISLMLIGSVLGLLLALVLC 1144
DB 1051 LSVFNSKSHRLVYSKASKAEQVLVKVLIHEKML-HVYVLSG--IGGLVILFLIFLAL 1107
QY 1145 WKLGFFAHKKIPEEKREKLE 1166
DB 1108 YKVGFF-----KNTLKEKME 1122

RESULT 5
Q9R200 PRELIMINARY; PRT; 1160 AA.
AC Q9R200;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Integrin alpha L.
GN ITGAL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SPLEEN;
 RA Ma R.Z., Teuscher C.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF065901; AAD25884.1; -
 DR HSSP; P20701; 1LFA.
 DR MGD; MGI:96606; Itgal.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 SQ SEQUENCE 1160 AA; 128127 MW; A33C531B139F1EAD CRC64;

Query Match 17.8%; Score 1089.5; DB 11; Length 1160;
 Best Local Similarity 28.3%; Pred. No. 2.4e-73;
 Matches 359; Conservative 189; Mismatches 454; Indels 265; Gaps 49;

12 PLVFLTGL-----CSPFLIDEHHPRLFPQPPAEFGSYLVQHVGGGQRMVLGAPWGDPS 66
 8 PRLLLGLQFLPAKWSYNDLTPQSLAQAGRHFGYQLIEDG---VVVGARGE--- 60
 67 GDRRDVRCFPGVGAHNAAPCAKGLGDYQVGNSSHPAVMHLGMLLEIDGGGMACAP 126
 61 GDNLTGLVHCRTSSEFCQPV-----LHGSMN--TSKYLGMTLADDAKSLACDP 110
 127 LMSRACGSSVFSSGICARVDASFOPGSLAPTORCPTY-----MDVIVLIGSNST- 178
 111 GLSRICDONTYISGLC-----YLPFQSLGEPMLQNRPAQYECMKKGVLDVLPDSQSLD 165
 179 -YPMSEVOTFLRLVGLFIDEQIQVGLVQVGGSPVHEWSLGFRTKEEVVRAKNSLR 237
 166 RKDFEKILEFMKDVMKTL--SNTSYQFAVGFSTDCRTEFTPLDY-----YKQKNBDV 217
 238 REGRE-----TKTQALIMVACTEGFSQSHGGRPEARLLVVTGESHIDGELPAALKA 291
 218 LIGSVQPMFLTLNTRAINVVAHAVFKESGARPAKTKVLIITGEMSDKNISA--- 273
 232 CEAGVTRYGIVAVLGHVLRORDPSSFLREIRTIISDPDERFFNVNTEALTDIVDALG 351
 274 --AHDTIRYIIIGIGHFVSVQKQ---KTLHIFASEPVEEYVKLIDTFEKLKDLFTDIQ 326
 352 DRIFGLESSHANESSFGLEMSQIGFSTHRLKDGLFGVNGAYDWGGSVLMLEGHRLFP 411
 327 RRIYVIEGTNRQDLTFMNEELSSGSIADLSKGHVVGVAGAKDWAGGFLDLR----- 379
 412 PRMALED-----EPFALONHAAYLIGYSVSMILRGGRFLFSGAPRFHRRGVIAF 463
 380 -----EDIGATFVGOEPLTSDVRGGYGYTVAMWTSRSRPLLAAGAPRYHVGQVLLF 434
 464 QL-KKDGAVRVQSLQGEQIGSYFSGELCLPLTDDTDVLLVAAPMLGPQNKETGRV 522
 435 QAPKAGKRNQTKIEGTQIGSYFSGELQSVLDDQGEHELILIGAPLPFGGQ--RGGRV 492
 523 VYVIVGQOSLLTLQGTLOPEPPQD--ARFGFAMGALFDLNDGFAVDVAVGAPLDEHQAL 581
 493 FTY--QRQGLFEMVSELQDGPYIGRFAALITALLTDINGDRILTVAAGAPLEB--QGANV 549
 582 YLYHGTQSGVRHPQRILAAASMPHALSTFGSGVDCRLLDDDDLVDAVAGQAAILIS 641
 550 YIFNKPGGSLSPQSQRIQGAQVFPDIRWFGRSIHGVKDLGDBRIANVAVGEGGVVVIS 609
 642 SAPIVLTLSLFTVTPAIVSVQDQ---RRRGOEAVCTTALCFQVTSRTGRMDHORY 697
 610 SRPVDVVTVELSFSPEIIVHVEGYSKAREQKGVKAKA--CRIRIKLTP-----QFO 662
 698 MR-----FTASLDWTAGARAAFDGSGGRLSPRLRLSVGNV-----TCBQLAHFV-L 744

Db 663 GRLLANISYTLQIDLGHMRNRSGLFPDGSHEL-----GNTSITPDKSLDFHFPPI 714
 QY 745 DTSYLRVALVTVFALDNTTKPGPVINEGSP-----STOKLVFESKD 788
 Db 715 CIQDLISINVSINFLS-----LEEETPRQOKGRAMOPILRPSHYTKETKIPFKN 766
 QY 789 GCPNCECVTLVQVNDIRSGRAPFVVRGRRKVLSTLENKENVATSTSIIFSR 848
 Db 767 CGEDKKCEANLTLS-----SPARGPLRIMS--ASLAVEMWTLNSGSDAYVVRDLDPFR 820
 QY 849 NLHLASLTP-QREPIKVECAAPSAHARL-----CSYGHVPQTGAQVTLLEPERSCS 901
 Db 821 GLSFRKTEMLOPHSRMPVSCSELTEGSSLLTKTLKCNVSSPIFRAQGEVS----- 870
 QY 902 SLISQVREKLTASS--PSLENGTL-----QENTACTSAVIOYEPHLFSSSESTL 949
 Db 871 --LQVMEFTLINSWEDVELNGVHCENENSSLOEDNSAATHLPVLYPNILTKEDN- 927
 QY 950 HRYEVHPYGLTPVGPGPEFKTLRVQNLGCYVSGLLIISALLPAVAGNYFLSLQVIT 1009
 Db 928 -----STLYISFTPRGPKTQGVHV-----YQVRIQPEAY 957
 QY 1010 NNASCIVQNLTPRPGPVNPEEL-----QHTNRLNGSNTOCVVRCGLQGLAKTEVS 1063
 Db 958 DHNPMTLEALGVDRP--HSEDLITYTWSVQTDPL-----VTCHSBDLKRPSSSEAE 1006
 QY 1064 GLRLVNHNEFFRRKPKF---SLTVSTFELCTE--EGSVLOLTERA----- 1103
 Db 1007 PCLPGV--QFRCPVFRWEIILIQVTVELSKETIKASTLSLGSLSVSNSSKHFLYG 1064
 QY 1104 SRWSES-----LLEVQTRPILISMLIGSVLGGILLALLVFCMLKGFPAHKIPREE 1159
 Db 1065 SKASEAQVLVAVDLIDHEKEML-HYVVLSG--IGGLVILFLFLALYKGF-----KR 1114
 QY 1160 KREEKLE 1166
 Db 1115 NLKEKME 1121

RESULT 6
 Q9OXH4 PRELIMINARY; PRT; 1169 AA.
 AC Q9OXH4;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Leukocyte adhesion glycoprotein p150.95 alpha integrin subunit.
 GN ITGAX.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Huang X., Gorski K., Tong C., Rattis F.-M., Tseng S.-Y., Pardoll D.,
 RA Tsuchiya H.;
 RT "Isolation of Genes Selectively Expressed by Dendritic Cells";
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF211864; AAF23492.1; -
 DR HSSP; P1215; IABX.
 DR MGD; MGI:96609; Itgax.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
 DR PROSITE; PS50234; VWFA; 1.

```

KW Integrin. 1169 AA; 129150 MW; C616412033C219A6 CRC64;
SQ SEQUENCE

Query Match 17.5%; Score 1069; DB 11; Length 1169;
Best Local Similarity 28.3%; Pred. No. 8.6e-72;
Matches 354; Conservative 213; Mismatches 490; Indels 196; Gaps 49;

QY 1 MELPFTNHLFLPLVFLTGCSPPNLDENHRLFPPEPEAFEGSYLVQHVGSGGRMMLVGA 60
DB 1 MCGTWAIFLL-LGFVS--CLGPNLDAEKLTHP-HNDGAEFGHSVLYQY---DSSWVVA 53
QY 61 PWDGSGDRRDYRCVPVGAHNAAPCAKGLGDYQLGNSSHPANNHLMGSLLETDGCG 120
DB 54 PREIKATNOIGGLYKC---GYHTGNCERISL-----QVPEAVNISLSTLSLAATNPMS 104
QY 121 FMACALMRACSSVFSSGICARVADSPQGSALPTAQR-CPTV-MDVVIVLDGNSI 178
DB 105 LLAAGTVAHTHCENIYLTGLCLFLSSSFK-QSQNPPTAQOECPODDIVFLIDSGSI 163
QY 179 --YPMSEVOTFLRLVGLKFLIDPEQIQVGLVQYGESPVHWSIGDPRTEKEEVRAKNLS 236
DB 164 SSTDFEKMIDFVKAWSQL--QRPSTRFSLMQSDYFRHFTNNNTISTSSPLSLGSR 221
QY 237 RREGRETQAQAIMVACTEGFSQSHGGRPEARLLVVTDSH-DGEELPAALKACEAG 295
DB 222 QLRG-YTYASAIKHVITELFTQSGARODATKVLIVITDGRKQDNLSYDSVIPAWEAA 280
QY 296 RYTRVGIANGLHRLRRORPSSFLREIRITASDPDRFFNFYDEALDIDVALGDRIF 355
DB 281 STIRAIAG-GKAFVNEHSK---QHLKAIASMPSEHYFVSENPALQDIENQLEKILF 335
QY 356 GLEGSHAENESSFGLMSOIGFSTHRLKDGILFGMVGAVDGWSVLMLEGHRLPEPMA 415
DB 336 AIBGTETPSSSTFELMSQGFSAVTPPGVIGANQGSW-----SGCAFVLYSNMR 388
QY 416 LDEFPALQNH-----ATLGYSVSSMLRGGRRLFLSGAPFRFRKXVIAFOLK 467
DB 389 -----PTFINMSQENEDMRDALYGTSLAFWKVHSLIL-GAPRHQHGKVIF-TOE 440
QY 468 DGAVRVAQSLQGEQISYFSGELCPDTRDGTDTVLVAAPFLPGPKEKRGVYV-L 526
DB 441 SHHWRKSEVRGQIQISYFGASLCSDVMDQSDTDLVLGVPRYV--ETRRGQSVSCM 498
QY 527 VGOOSLLTQGLQEPPOD-ARFGAMGALPDLNDGFAVDVAVGALPDGHALYLVH 585
DB 499 PGVGSWMHGCTTLHGGHPWGRFGAALTVLGDVNDLSADVAIGAPGEENRGAUYIHH 558
QY 586 G-TQSGVRPHAPORIAAASMPHALSTFGASVDRGLDLDGDDVLVDVAVGQAAILLSRP 644
DB 559 GASRODIAFPSQRIASQIPSRIOYFGQSLSGGQDLTRDGLVDLAVGSKGRVLLTRTP 618
QY 645 IYHLFSLLEVTPOAISVQDRCRR--GOEAVCLTALACQV-TSRPRGMDHQFYVRFTA 702
DB 619 IIRVSTVFHTPRAEISRSRVPECOEQAIPQOTLSDAIVCHIHESPQTQGLDKRSTVFDL 678
QY 703 SLDEWTAGARAAFDGSGQRSLPRRLSLVGNVTCEQDHFHV-LDTSILYRPVALTVTFAL 761
DB 679 ALDHGRLSTRALFKETKTRALTRVKTLGL-NKHCESVKLLLPACVDSVTPILRLNFSL 737
QY 762 ----DNTTKPGVNLNESPSTSIQKLVPSFKQCGPNECVTDLVLQVNDIRGSKRAPV 816
DB 738 VGVPISSLQNLQMLLAVDDQYFTASLPFEKNGADHICODDLISV-----VGFPPDLKLT 792
QY 817 VVGGRKXVIVSTLLENKENAVNTSLTIF-----SR 848
DB 793 VVGSDELAVNDVVSNDGEDSYGTITLYFPVGLSFRVAEGVFLRKDEQDQMRGRGH 852
QY 849 NLHL-ASLTPORESPIKVECAAPSAHARLCSVGHPIFOTGAKTFLLEFSCSILSVOY 907
DB 853 SLHLMQDSTFPRDSQGL-----WSTSCSSNHVIFRGGSQMTFLVTVDVGPXKMLGR 903
QY 908 FGLKLNASLSLEBN-GTLQENTAQTSAYIOYEPHLLFSSESTL----- 949

DB 904 L--LIRARVGSBNVPGT-PKTTFOLELPVKAAYVTMISSHDQTKYLNFTSEKEXTSV 960
QY 950 --HREYVHPYGLTPVGPGEFFKTLRVQNLGCVYVSGLIISALLPAVAGNVFLSLSOV 1007
DB 961 VEHRFOVNNLGRDVPVAINFWPIELKGEAVWVY-----MWHPON--PLQOC 1007
QY 1008 ITNNASCIQVNLTEPPGPVHPEELOHTNR--LNGSNTQCVYRCHLGQALAKTEVSVG 1064
DB 1008 YRNRLK-----PTQFDLLTHMQKSPVLDCSIADCLHLRCDI-----PSLG 1047
QY 1065 LIRLVHNEFFRAKSLTIVSTFELGTEEGSVLOLTA-----SRMS-----ESLLEV 1113
DB 1048 ILDELY--FILKGLNLSFGWISQTL-----QKQVLLSEAEITNTNTSVYSQLPGGEAFIR- 1099
QY 1114 VQTRPIL-----ISLWILIGSVLGLLILLVLCWKLGF--AHKIPPE 1158
DB 1100 AOTKIVLEMYKVNHPVPLIVGSSVQGLLLLAITITAILYKAGFFKQYKEMLEE 1152

RESULT 7
09J130 PRELIMINARY; PRT; 1151 AA.
AC 09J130;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Integrin beta 2 alpha subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RA SEQUENCE FROM N.A.
RP Fathallah D.M. Sr.; Zeria K. Jr.;
RT "Cloning of the rat CD11b cDNA sequence.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF268593; AAF81280.1; -.
DR HSSP; P11215; IABX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF01839; FG-GAP; 5_A.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 1151 AA; 126943 MW; 8F785695D4074CAs CRC64;

Query Match 16.4%; Score 1000.5; DB 11; Length 1151;
Best Local Similarity 28.4%; Pred. No. 1.3e-66;
Matches 346; Conservative 216; Mismatches 491; Indels 165; Gaps 45;

QY 19 LCPSPNLDENHRLFPPEPEAFEGSYLVQHVGSGGRMMLVGAAMDGSGDRRDYRCPV 78
DB 13 LCHGNLDTENPMFQ-ENASGFGSYQL---GETRVVAAPQEVKAVNQTGALYQC-- 66
QY 79 GGAHNAAPCAKHLGDYQLGNSSHP-----AVNMILGMSLLETDDGGMACAPLMSR 130
DB 67 -----DYST-NRCDPIPLQVPEAVMMSLGLSLAATVTPPOLACGPTVHQ. 111
QY 111 AGGSSVFSSGICARVADSPQGSALPTAQR-CPTV-MDVVIVLDGNSI--YPMSEVOT 186
DB 112 NCKEYTVNGLCYVFGSLRLKPPQGFPAALGCPQGESINAFLLDGGSSINTTIDFQYKE 171
QY 187 FLRLVGLKFLIDPEQIQVGLVQYGESPVHWSIGDPRTEKEEVRAAKNLSRREGRETXTA 246
DB 112 FVSTYMDQ--FQKSTLFSLMQYSDPEPTHFTFNDFKKNPDKSHVRIQLNGR-TKTA 228
QY 247 QAIMVACTEGFSQSHGGRPEARLLVVTDSHDEGL--PAALKACEAGRYTRYIA 303

```

Db 229 SCIRKAVRELFQKINGARDNAKILVITIDGKFF-GDPLNTYDVLPPEAEAG-TIRYIIG 286
 Qy 304 VLGHVLRQRDPSSFLREIRITIASPDERFFNVITDEALITDIALGDRIFGEGSHAE 363
 Db 287 V-GNAFHK---PQS-RRELDITASKRAGDHVQOVNFELNITRNQLOEKIATIGSTQIG 341
 Qy 364 NESSFGELESGQIFGFTHRLKDGILGSMGAYDM-GGSVLMEGHRLPPRMALDEDEPP 422
 Db 342 STSFEHEMSQEGFSAITNSGPLGSGVSPDMAGAFILPSKDXASFTNTTRIDSDM- 399
 Qy 423 ALQNHAAVLGYSVSSMLLGGRRFLSGAPRRHKGXVAFOLKXDGARVAOSIQEGOI 482
 Db 400 ---NDATLGTIA-SAVISRNVOVLGAPRYOHLGLVWFK-QNFGAHEPHTDIKSGOI 453
 Qy 483 GSYFSSSELCPDTRDGTDLVLAAPMFLGPNKTEGRVYVYLVGQ---QSLITLQIG 537
 Db 454 GSYFASLCSVMDADGNTNLLIGAPHNY--EKTRGGGVSCPLPRGARQCCAILHG 511
 Qy 538 TLQPEPPDARFGFAMGALPDINOGFADAVAGARLEDHGOALYLVHGTQ--SGVRPH 595
 Db 512 D-QGHP--WGRFGALTLVGDVNGDKLIDVAIGAGEOEONGAVYIFHGASVASISTPH- 567
 Qy 596 AORIAASMPHALSYFGRSVDRDLDDDLVDVAVAGQAMILLSSRPVHILTSLEVT 655
 Db 568 SORIGARPSGLOIYFGOSLSGKDLTMDGLMDLAVAGQGRILLRAQEVRLKATMEPS 627
 Qy 656 FOAISVORDCRRR-----GQEAVCUTLAALCFVYTSRTPGR--WDHQFYMRFTASID 705
 Db 628 PGNLARSVACQEEVYVKNKDAEYVCL-----QVRKTKRLKLEBGDIQSTVYTDLALD 681
 Qy 706 EWTAGARAAFDGSGGRLSPRLRLSVGNVTCQOLHFHVLDTSD-VLRPAALVTTPA--- 760
 Db 682 FGRSVYRAFFEDTKNGIL-RRIRVFGLTQKCELTKLILPDVDNSVSPILRLNTLVGE 740
 Qy 761 -LDNTTQCPVLNBSPTSIOKLVPESKDCGPDNCTVDVLQVNMMDIRGSKAPFVVG 819
 Db 741 PRSSRDRLRPVLAEMAQRIETAMPPEKXGMDTICDDLSITVS---STRVNTLVVG 795
 Qy 820 GRRKVLVSTTLERKRNAYNTLSIIFSRHLH---ASLTPORESPIKECAAPSA--- 872
 Db 796 DSRDQDVSTLNRNDEDSGTGKTCYBGLSYRKVASQNOFSKPKMRV-IAEFSSESG 854
 Qy 873 ---HARLCSVGHVFOGAKVTPLEEF--SCSSILLSQVFGKLTASDSIERNTLOEN 927
 Db 855 QGVLTSTIMDINHPIFPANSEVTFPVTFDVDSADSLNKLKLVVTS---ENNVSRDIX 911
 Qy 928 T-AQTSAYIQYEPHLLFS-ESTL-----HRYEHPHG--TLPIGPGPE 967
 Db 912 TESQLELPVKATVYVVTSGESSIKYLNFTASEMTSKYIOHQYQFNMLGQRSFPVS--- 967
 Qy 968 FKTLRVONLGCYVSGILISALLPAVAHGNYFSLSOVITNNASCIVONLTEPPGPV 1027
 Db 968 -----VFWIIVQINKVTIMPPPOVTSQONLSSVCRTQKSPSSK 1008
 Qy 1028 HPEELQHTNRNLNGSTQCVAVCHLGOLAKGTESVGLRLVHNEFFRRAKKSLLTVST 1087
 Db 1009 FODELERPVNLCSVAVCKRIQCDIPSFNSKEIFNVTLQGNLLFPWYIETSHDLVST 1068
 Qy 1088 FELGTEEGSVLOLTASKRSESLLEV-VOTRPIILSLMILISVGLLALLLVFCIMK 1146
 Db 1069 AEILNDSAFALLPQOETFEVKQETETKVEPYVHNVPPLIVGSSVGVLLALLITAGLYK 1128
 Qy 1147 LGFFAKKKIPEEKREEK 1164
 Db 1129 LGFF-----KQYK 1137
 RESULT 8
 ID 006271 PRELIMINARY: PRT: 780 AA.
 AC 006271;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Integrin alpha-2 subunit (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 NCBI_TaxID=8355;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Meng F., Desimone D.W.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 95-168 FROM N.A.
 RX MEDLINE=94008528; PubMed=8404528;
 RA Whitaker C.A., Desimone D.W.;
 RT "Integrin alpha subunit mRNAs are differentially expressed in early
 Xenopus embryos.";
 RL Development 117:1239-1249(1993).
 DR EMBL; L43058; AA69770.1; -;
 DR EMBL; L10186; AA16246.1; -;
 DR HSSP; P11215; 1A8X.
 DR InterPro; IPR000413; Integrin_alpha.
 DR Pfam; PF01839; FG-GAP; 4.
 DR Pfam; PF00357; Integrin_A; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR SMART; SM00191; Int_alpha; 4.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 FT NON TER
 SQ SEQUENCE 780 AA; 87017 MW; 9518B18C2B66F637 CRC64;
 Query Match 16.3%; Score 997; DB 13; Length 780;
 Best Local Similarity 31.6%; Pred. No. 1.3e-66;
 Matches 253; Conservative 161; Mismatches 338; Indels 48; Gaps 17;
 Qy 391 VGADWGSVUMLEGGHRLPPRMALDEPPALQNHAAVIGYSSMLLGGRRFLSG 450
 Db 1 VGADWGSTVTHESKQSFIPYKAFKVLND--RNOSYLGYSVAAYNMNVH- 57
 Qy 451 APRFRHKGXVAFOLKXDGARVAOSIQEGISGFSSELCPDTRDGTDLVLAAPM 510
 Db 58 APRDITGQVYVYVITSGANSVIOTQREGDISYFGSVLCSVNVNRDSTIDVLVGAFT 117
 Qy 511 FLGPNKTEGRVYVYLVGQOSLTLQGTLOPEPPDARFGFAMGALPDINODGADYAVG 570
 Db 118 FMNEYKKEBGQVWFSTRIDGLVQREGPKSLNTRFGSAIYELSDIDGNDVIYVG 177
 Qy 571 APLEDHGOALYLVHGTQSGVRPHPAQRIAAASMPHALSYFGRSVDRDLDDGDLVDVA 630
 Db 178 APLENONGSAIYIPNGKNITRTYXSKILGSSFNPGLOYFGRSVGDHODINGDTINDVS 237
 Qy 631 VGAOGAAILSSRPVHILTPSLEVTPOAISVORDCRRRQGEANCLTAALCFVYTSRTPG 690
 Db 238 VGATGKIYQLMSQVADVAVKALFIPOQIVLTNKA------VTRICFSAVFR-PA 288
 Qy 691 RMDHQFYMRFTASIDWTAGARAAFDGSGGRLSPRLRLSV---GNVTCQOLHFHVLDTSD 747
 Db 289 NSQSVAVTYATLADADLLSRVTSRQGFRRSNDRFIQNRNIVIGPTSCFHHVNVQETS 348
 Qy 748 DYLRPAALVTVPALDNTTKG--PVLNBSPTSIOKLVPESKDCGPDNCTVDVLQVNM 805
 Db 349 DSENALRLRVNVAQN--PSSSPVLNPPSSSTEWPIPLKDCGDCKICISDSLQ-Q 404
 Qy 806 DIRSGRAPFVVRGGRKVLVSTTLERKRNAYNTLSIIFSRHLASLTPORESPIKY 865
 Db 405 QIPSDIKSPYVSKMDRRLFOVLTNKLKVAAYTRNALFSENLFPASSTIPPDG-TEV 463
 Qy 866 ECAAPSAAHL-CSVGHPVFQGAKVTFLEFEFSCSLSOVFGKLTASDSIERNTL 924
 Db 464 LCQVGNQNSVSCILIGFPLKDSOKVFDLWFDNANLNTAIFYQATISES--REARE 521
 Qy 925 QENTAQTSAYIQYEPHLLFSSESTLHRYEVHPTGLFV-----GCPPEFTTLRVONLG 978
 Db 522 EDNSVNTLTPVOYDAEHLTRVTNINFEYEVFPGINIPSVVNTPDIDGLEFNTVYK- 580

QY 979 CYYVSGLLISALLPAVAHGNYFLSLOYITNNASCIVONLTBPG-----PPVHPD 1030
DB 581 TYPPLKAYNISINVPOTKSNPLIYITSVOTDQMTGVSSNRVNPAPHVEKIPKBPSEK 640
QY 1031 ELQHTNRNGSNTOCCQVVRCHLQOLAKGTEVSVGLRLVHNEFFRAKESLTVSTFEL 1090
DB 641 NIKAVHVNLCESARCAVAKICFLENLQUREDYVNVSTRIMNGTFAAHPQTELVANAKL 700
QY 1091 GREBSVLOLTEASRMSSESLLEVOTRP---IISIMILIGSVLGGLLILALLVFCMLK 1147
DB 701 ETIDSEIFVYVSN---KDWMIPLITRKPEKSEVPVIGVISAIGLVALLALVALLMLK 756
QY 1148 GFFAHKKIPBEEKREEKLEQ 1167
DB 757 GFF-KRKYOKLQTEEBEIRE 775

RESULT 9
Q96HB1

PRELIMINARY; PRT; 1086 AA.
ID Q96HB1; AC Q96HB1; DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Unknown (protein for MGC:1714).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
RN (1) SEQUENCE FROM N.A.
RP TISSUE=LYMPH;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008777; AA008777.1; -
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; EG-GAP; 5
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; Vwa; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
DR PROSITE; PS0234; VWF_A; 1.
DR DR PROSITE; PS0234; VWF_A; 1.
SQ SEQUENCE 1086 AA; 119223 MW; F6F2546B8C32F9 CRC64;

Query Match 16.1%; Score 984; DB 4; Length 1086;
Best Local Similarity 27.7%; Pred. No. 2.1e-65;
Matches 344; Conservative 179; Mismatches 427; Indels 290; Gaps 51;

QY 11 LPLVFLTGL-----CSPNLDENHPRLLFCGRPEA--EFGYSVLQHYGGQRMVLVGA PMD 63
DB .9 MAMALLSGFFFPAPASNYLDVRCASFS--SPRAAGHFGYRLVQ--VGNG---VIYAPBE 63
QY 64 GSGDRGDRGVRCVPVGAHNAAPCAKHLGDYOLGNSHPA VNMHLGMSLLETDDGGFMA 123
DB 64 GNS---TGSLYOCQSGSTGCHLPVT-----LRGSNY--TSKYLLMTLATPTDGSILF 110
QY 124 CAPIMRACGSSVFSSGICARVDASFQPGSLAFTQRCPTVMDYIVIVDGSNSIYPMSE 183
DB 111 AA----- 112
QY 184 VQFLRLVGLFLIDPEQIQVGLVOYGESPVHMSLGF---RTKEVVRAAKNLRRREG 240
DB 113 -----VQSTSYKTEFPDSDYVKRDPALLKHYKHMILL-- 147
QY 241 RETKTAQAIMVACTEGFSQSHGGRPEARLLVVTDDSGHDEELPAALKACEAGVTRY 300
DB 148 --TNTGAIINYATEVEFRELGA RPDATKVLIIITDGEATDSCNDAAD-----IRY 199
QY 301 GAVLGHYLRQRDPSPFLREIRTIASDPDERFFPNVTDEALTDIVDALGRIRFLEBS 360
DB 200 IIGIGHNFTKESQET-----LHKFASKASEPFVKILDTPEFKKLDLFTLEKKIYIEGT 254

QY 361 HAENESSFGLMSQIGSFTHRLKODILFGMYGAYDMGSGVLMLEGHRLFPFRMALEDEF 420
DB 255 SKQDITSTNMELSSSGISADLSRGHAAVGAAGAKMAGGFLDLKADLQ-----DDTF 306
QY 421 ---PPALQNHAAVYIGYVSSMLRGRRLLPLSGAPRRFRHGRKYAFOLKDDGA--VRYAQ 475
DB 307 IGNEPLTEFVREAGYLGIVTVMPSRKQKTSLSASGAPRYOHMGRLVLPQEPQGGHMSOVQ 366
QY 476 SLQGEQISYFSGSELCPLDTRDGTDLVLAAPFLGEPQNKETGRVYV---LVGQOSL 532
DB 367 TTHGTQISYFEGELCGVDVQDDEFTLLLGALFLYGEQ--RGGRVFTYORRQLEFBEV 424
QY 533 LTLQGLTQPEPPQDARFGFAMGALPDNLQDGFADVAAPALDEHQALYLYHGTOSVVR 592
DB 425 SELQG--DPGYPL--GRFEBATLALTDINGDGLVNDVAVAPLEE--QNAVYIFNGHGLS 479
QY 593 PHPAQRIAASVPHALSYFGRSVDGRDLDDGDDLVDAVGAQGAAILLSRPYHLLTSL 652
DB 480 FQPSQRIEGTQVLSGIGWFRSIGHVKDLEGDLADVAVGAESQMIIVLSSPPVDMVTLM 539
QY 653 EYTPQALISVQDRC---RRGQEA VCLTALCFQVTSRTGRMHOQVYMR-----FTA 702
DB 540 SFSPELIPVHEVEGYSYTSNKKKEGVNIT--ICFQIKSLIP---QFQRLVANLYTL 592
QY 703 SLDEWTAAGARAFDSSGQRLSPRLRLSVGNVTCQOLHFHY--LDTSDYLRPVALTVPAL 761
DB 593 QLDGHTRRRGIFPGGRHEL--RNIAVATMSCTDSEFHPVVCQDILSPINSLNSL 650
QY 762 -----DNTTKRGPVLANGSPSTIOKLVFSDCCGPDNECVTDLVLOVN--MDIRSR 811
DB 651 WEEEGTPRDORAGKIDPILRPSLHSETWEIPEFKNCEDEKCEANLRLVSSFPARSRLR 710
QY 812 KAPFVVRGRRKVLVSTLENKKNAYVNTSIIIFSRNHLASLTPOR--ESPIVEKCAAP 870
DB 711 LTAFA-----SLSVELSLNUEBDAYVQDLHLRPPCLSRKVMELKPHSQIPVSGCEL 764
QY 871 SAHARL-----CSGVHVPQGTGAKVFTLFEFFSCSILSQVFGKLTASS--DSLERNG 922
DB 765 PEESRLRLRALSQVNSPIFKAGHVA-----LQMMFPTLVNSMGDSVELHA 812
QY 923 T-----LQENTAOQTSAYIOYEPHLF--SSESTLHRYEVNRPVGTLLVPGPPEKKT 970
DB 813 NVTCKNEDSDLLIEDASATIIPIILVPIINILQODEDSTLY-----VSFTKRG----- 860
QY 971 TLRVGNLCYVYSGIISALLPAVAHGNYFLSLOYITNNASCIVONLTBPG--GP--- 1025
DB 861 -----KIHQVKHYVQRIOPSI-HDHN-----ITLEA--VVGVPDPSPSGPRTTH 902
QY 1026 -----PVHPELOHTNRNGSNTOC---QVVRCHLQOLAKGTEVSVGLRLV--- 1069
DB 903 QMSVGMERPVPCHVEDLE---RLPDAAPCLPGLALFRCPVAFROELIVQVIGTLEVGEI 959
QY 1070 -HNEFFRAKESLTVST--FELCTEGSVLQLTASRMSSESLLEVQTPILISMLIL 1126
DB 960 EASSWFSLCSSLSISIFNSKHFHLYGNSASLAOVV-----MKVDVYERKOMLY-LVVL 1011
QY 1127 IGSVGLGLLILALLVFCMLKGFPAHKKIPBEEKREEKLE 1166
DB 1012 SG--IGGLLLILFIYLVYKGVF-----KRLKEKME 1042

RESULT 10
Q98TF1

PRELIMINARY; PRT; 1196 AA.
ID Q98TF1; AC Q98TF1; DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CD11-1.
CN C1A1.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;

OC Cyprinidae; Cyprinus.
 OX NCBI_TaxId=7962;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PERITONEAL EXUDATE CELLS;
 RA Kimura M., Nakao M., Miura C., Fujiki K., Yano T.;
 RI "Molecular cloning of a leukocyte integrin from the common carp."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB048536; BAB39134.1; -
 DR HSSP; P20701; ILFA.
 DR InterPro; IPR001699; Asparticase site.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFDOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE; PS50234; VWF; 1.
 DR PROSITE; PS50234; VWF; 1.
 SQ SEQUENCE 1196 AA; 132477 MW; 9369C807E7DCA53B CRC64;

Query Match 16.1%; Score 980.5; DB 13; Length 1196;
 Best Local Similarity 27.1%; Pred. No. 4.7e-65;
 Matches 334; Conservative 213; Mismatches 526; Indels 159; Gaps 48;

QY 5 FVTH-----LFLPLVFLTGLCSFNDENHPLFPFPAPAEFGSYLQHVGGGRMM 56
 DB 6 FVTQTSSRMETLGLFLFWASLSEAFNIDTEHPLRFNGAPEDFGYSYOTEFGNRCOI 65
 QY 57 LVGAWPDGSGDRDGVRCPVGGAHNAFCAGKHLGDYQGNSSHPAVNMHLSLETD 116
 DB 66 IVGALLEGNS---TEEMVST---ADLQSCQR-----LQPPGESVRF-FGMAAAS- 110
 QY 117 GDGFMACAPLWSRACGSSVFSSGICARVDASFQDGSIAPTAQR-PTMYDVIVLDGS 175
 DB 111 -SAALTSCSPYPHEDCDNSYLVNGVCYFSSLSQVNSFTAYQSCSREVLVLFEDS 169
 QY 116 NSI--YPMSEVQTFILRLVGLFIDPEQVGLVQGESPVHEMSLGRPT---KEVVR 230
 DB 170 SSMKTVDDEMNKNFTKIDIMKKL--SNSSIKFAVQFSTDVTFVFPNDVQSSGAEEKMK 227
 QY 221 AA-KNLRRREGREKTAQAIWVACTEGF-SQSHGGRPREARLLVVTVDGESHDEBLPA 287
 DB 228 ETHMKSLS-----INTKAIYIILKNLNSMLSGADSKAQALVIITDGDSDMDYD-N 279
 QY 288 ALKACEAGRVTRYGIAGVGLHVLRRQRPSSFLREIRTIASDPDERFFFNVTDEALTDIV 347
 DB 280 VLKCCDEQIILKYYIGV-----GKVDLIELTQLASEPKRNTFYIKYSGKLGL 329
 QY 348 DALGRTIGLESHAHNESSFGLMSQIGFSTHRLKDGILFGMVGAYDWGSSVLMLEGH 407
 DB 330 DNLQKKINIEGSSVAQGRDROKELISQSFVSVYOEESVIVGVSNDMRGALYVTSG 389
 QY 408 RLFPRLMLDEFPALONHAAYVIGSVSMILRGGRRLFGSGAPRFRHGVIAFOLKK 467
 DB 330 SDFK-----ETIILPVAVKDSYMGYS-TVLKMRGVSILFSGAPAEHTGIVTLF-TKN 442
 QY 468 DGAVVAOSLQEQIGSYFGESELCPILDRDGTDTVLVLAABMFLGPONKETGRVYVVLV 527
 DB 443 QNTWTVTSNINQEQIGSYFGASLILDDVSDSDSDFLVGALFLFQSGQRTGRLVYVSL 502
 QY 528 GQGSLLITLQGTIQEPDQARFGFAMGALPDLNQGAFADVANGALEPDECHQALVYHGT 587
 DB 503 SQOKY-FQKTLNVSQSTTGRPAASVASISKLDINGGLSVAAVGALE-NEGIVVYVIGD 558
 QY 588 OS-GVPRPH-PAORIAAAMPHALSYFGRSVDRGLDLDGDLVDVVAQAGAILLSSRPI 645
 DB 559 RHGHNELTLPORISVQSVLPQLQDQVSLTQGMNNDNLTDIYGAQGGVLLKARV 618
 QY 646 VHLTSLLEVTPQAISSVQDRCCRQGEAV-CLTALCFQVTSRT--PGRMHDQFYMRFTA 702

DB 619 MSVSAQLSFSPEKISLNFEC--PGSNAFNAFNLTSCPTVTERSTSTSLKKLVNSLNL 676
 QY 703 SLDEWTACARAAPGSGGQRLSPRRRLSLV---GNVTCEQLHFHVLID-TSDYLRPALVTV 758
 DB 677 NVDVYRGMSRGFPFQDS--VSSRTIQGSVLDDSGSSCNFSFIPIRLCAVDVSPFLKIMN 734
 QY 759 PALDNTTKPG--PVYNEGSPYSIOKLVFYSXDCGPDNCEVTDVLVQ--VNMDIRGSRKA 813
 DB 725 FSQTEMLSGNVAVILDVHSRTEENVEVPFGNCNNSNSCVADLKLNFSFTNNTLVENQA 794
 QY 814 PFVYRGGRKVLVSTLENKRENAVNTSLIIFSRNLHLASTTPQRESPIKVECA--AP 870
 DB 795 HFTVQ-----VSLNPPGDDSYNTSIVLHYEGISLSKFDKIKPRTSSCCGRDSC 845
 QY 871 SAHARLCSGVHPVQGTAKVTFLEFEF-----SCSSLSLYQFGLTASDSLENGTQOE 926
 DB 846 ATRRTTCSIDLPVRSRGTGTTQFLGTFPRVKKNDNSNMEIM--ITANSND--NGNRSD 900
 QY 927 NTAQTSAYIQYEPHLPFS---SESTLHRYEVHPYGLPVGPGEFTKTLRQNLGCVV 982
 DB 901 TEVRSRVVQPAVDLAISLVAEDSVTVMNPFLEBRGPRPL-----NTYKVENSGRDL 954
 QY 983 SGLIISALLPA-----VAGGNVFLSLQVYITNASCIVON--LTBPPG 1024
 DB 955 P-VSVTLTLPOTPRVILTPHTFSWHEVHH--SFISYHQIIT--MCLINKHLFFSP- 1006
 QY 1025 PPMHPEELQHNRLNGSTQGVVRCHLQGLAKTEVSGV--LRLVH-NF----- 1073
 DB 1007 -----ELSAVOOVRSSTGWSLREVECSQFDLNKSAVFNLTADARLONVKEYSFY 1060
 QY 1074 -FRAKFSKLTIVSTFELGTEGSLQVTEARSSESILEVQTRPILISIMILGSVIG 1132
 DB 1061 EFRKONVSIISALNLYNTSLNQTSSELKYNPHSQTEVKYEFVPPSLMLIVCGAV-G 1119
 QY 1133 GLLLIALIVFLMKLGFPAHKKIPDEEKREK 1164
 DB 1120 GFFFLIILFLFLKCGFFKRNR-PDEFVQEDE 1150

RESULT 11
 ID 088340 PRELIMINARY; PRT; 1167 AA.
 AC 088340;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Integrin alpha E1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98056820; PubMed=9394838;
 RA Brenan M., Rees D.U.G.;
 RT "Sequence analysis of rat integrin alpha E1 and alpha E2 subunits:
 RT tissue expression reveals phenotypic similarities between
 RT intraepithelial lymphocytes and dendritic cells in lymph.";
 RL Eur. J. Immunol. 27:3070-3079(1997).
 DR EMBL; AF020045; AAC23662.1; -
 DR HSSP; P11215; IABX.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 4.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFDOMAIN.
 DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWF; 1.

SQ SEQUENCE 1167 AA; 128970 MM; D88A2C38ACDC2AAE CRC64;
 Query Match 15.6%; Score 953; DB 11; Length 1167;
 Best Local Similarity 28.1%; Pred. No. 5.5e-63;
 Matches 367; Conservative 195; Mismatches 442; Indels 302; Gaps 65;

QY 14 VPLTGLC-----SPFNLDENH---PRLPPEPEAEFGYSVLQHV--GGGGRMLVGP 61
 DB 4 LPTFTLLCMASLKQGA FNMVDVMAWTALQPGAPAV---LSSLHLHDPSSNQCCLIVAR- 59
 QY 62 WDGPSSDRGDVYRCVQ--GAHNAFCAR--GHL---GDYQGNSSHPANMHLGMSLLET 115
 DB 60 --RSSRNRNTALYRCASISIPDEIACQVPEHICMPGRYQ-----GVTLV--- 102
 QY 116 DGGGFMACAPLMSR-----AC-----MDVIVLDGNSIYPMSEVO----- 140
 DB 103 GHNHGVLVICIQVQARFRSLNSELTGACSLTLPNLDLOQAVFSDLEGLDPCAHVDSGD 162
 QY 141 ICARVDASFOPOGSLAPTAQRCPTV-----MDVIVLDGNSIYPMSEVO----- 185
 DB 163 YCRSKGGS---TGEKKASARRRTVEEDEDGTETALVDGSGSIEF--SDPQAKNFI 218
 QY 186 -TFLRLVGLFLDPEQIOVGLVQGESPVHEKSLGDFR-----TKEEVRAAKNLSRR 238
 DB 219 STMRNPFYKCF---ECNFALVQYQAVIQTEFDLOESRDINASLAKVSIQVKEV--- 271
 QY 239 EGRETQAQAIWACTEGFSQSHGRPEARLILVWTDGESH--DGEELPAALKACAGRV 297
 DB 272 ----TITASMQHVLDNITIPSRGSKKALKWVVLTDGIRFDPPLNLTVINSPEMOCV 327
 QY 298 TRYGVAVLGHYLRQRDPSSFLNRTIASDPERFEFNVTDMAULTDIVDALGRIFGL 357
 DB 328 VAFALGV--GDAPFN---NNTYRELKLIASDPKEATFKVTNYSALDGLSKLOQIVAM 382
 QY 358 ESSHANESSFGLMSQIGFSTRHLKDQ--ILRGMVATYQ--GGSVIM--LEGHRLPFRPM 414
 DB 383 EGTVGD---ALQYQLAQTFSAQIIDKQGVLLGTVAQFVMSGALYSTRONGGCGFLNQT 439
 QY 415 ALDEDEPPALQNHAAVLGYSSVSMRLGRRLFLSGAPRRHGRKTIATOLKKDGAVRVA 474
 DB 440 AKED-----SRTOYSLTGLSL--AVLKAHGISIVAAQPHKLRGAV--TELKREDEEDA 492
 QY 475 --QSLGEOIGSYFSGELCPDLTDRDGTTLVLAAPMELGPONKTEGRVYVYLVOQ--S 531
 DB 493 FVRRIEGEMGYSFGSVLCRPVIDMDGTTDFLVAAPFY--HIRGEGRVYVYQVEQDA 550
 QY 532 LTLTQGLTQPEPP--QDARFGAMGALPDLNODGFALVAVGAPLE-----DG--HQAALYLY 584
 DB 551 SPSLAHTLSGHPGLTNRSGFMAAAGDINODKFTDVAIGAPLEGAGDAGSYSGSVITY 610
 QY 585 HGTSQSVRPHAPORIAAASMPHALSYFGRSVDRGLDLDGDLVDVAVGAQGAAILLSRP 644
 DB 611 NGHSGGLYSPSQIITAASSVASGLHFGMSVSGGLDPNDDGLADITVGSRDAVLRSP 670
 QY 645 IVHLTSLSEVTPOAISV--ORDCRRQOEAVCLTALCFQVTSRT---PGRMDHOFY 697
 DB 671 VVDLTVSMFTPDALPMVFIGKMDV-----LCFEVDSVVAASEPOL--REWF 716
 QY 698 MRFASLDEMTAGARAFDGSQORLSPRLRLSVGNVTEQOLHFHYLDT-----SDLYRP 752
 DB 717 LNFETVDVDTKORRQLQCEBSSGCCCLR--KMGSGFLE--HFWLISIEELCEBECFEN 773
 QY 753 VALTVTFALDNT---TKRGPVLNE--GSPTSIOKLVPSFGDCGPDCECTDVLQVNMNI 807
 DB 774 ITIKTYEFQTSGRADYDNPFTLDHAKESALFOL--PYEKDCKNKKFCIAETQLTNNISQ 832
 QY 808 RSRKAPFVVGGRKVLVSTLENKENAANTYSLSIIFSRNLHLASLTPORESPIK--V 865
 DB 833 QR-----LVGVTKETVNINISLTNGEDSYMTNMAALNPRNLQPKI-----QKPVSPV 882
 QY 866 ECAPSAHARL-----CSVHPVFO---TGAKYTFLLFEFFSGSLISQVFGKLTASSDGL 918
 DB 883 QCDDEKPVASVLMNCKIGHPIIKRSSVAVSVTWOLE-----ESVFPNRADITVT 933

QY 919 ERNGTLOENTAOISAVIQEPEHLLFSSESTLHREYVHPYGLPVGPEDEFTLLEVQNLG 978
 DB 934 ISNSN--EKSILARETSLQFRAFI-----AVLSRPV--MYMTSOSPDSHKEFFNV----- 983
 QY 979 CVVSGLLIISALLPAVAGNVFLSLQVITNNASCI-----VQNLTEPPGPV 1027
 DB 984 -----HGENLFGAVPOL-----QICVPIKLQDFQIVRVKULTKQD--- 1019
 QY 1028 HPEELQHTNRLNGSTQOVYRCH-----LGQLAKGTESVQ-----LLR-----L 1068
 DB 1020 HTECTQSEPPAGSDPVQVHEKMSVCAITSNKENVVAIVASIGHTKQLRDVSELPT 1079
 QY 1069 VANEFFRAKFKSL-----TVSTFELTEGVSVLQUTEASRMSSESLLEVOTRPLIS 1122
 DB 1080 LGEISFNKSIYEGNAEHRTKITIYFLKEE-----TR-----S 1114
 QY 1123 LMIIGSVLGGLLALLVFCIMWKLGFAHK--KIPBEKEKUE 1166
 DB 1135 LPLIGSSIGGLVVLVIALIFKCGFPRKRYQQLNLESTRRAQLK 1160

RESULT 12
 Q98TF0 PRELIMINARY; PRT; 1187 AA.
 AC Q98TF0;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE CD11-2.
 GN CIA2.
 OS Cyprinus carpio (Common carp).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
 OC Cyprinidae; Cyprinus.
 RX NCBI_TaxID=7962;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PERITONEAL EXUDATE CELLS;
 RA Kimura M., Fujiki K., Nakao M.;
 RT "Molecular cloning of a leukocyte integrin from the common carp."
 RL Submitted (Sep-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB048537; BAB39135.1; -
 DR HSSP; P20701; ILPA.
 DR InterPro; IPR001969; Asprotease site.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; Fg-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00193; Int_alpha; 5.
 DR SMART; SM00327; vwa; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1187 AA; 131778 MM; 85EDC7CA8B6B1C59 CRC64;

Query Match 15.5%; Score 947.5; DB 13; Length 1187;
 Best Local Similarity 26.7%; Pred. No. 1.5e-62;
 Matches 333; Conservative 209; Mismatches 504; Indels 199; Gaps 50;

QY 5 FVTH-----LPLVFLTGLCSFPLNDEHRLPFPPEPEAEFGYSVLQHVGGGGRMM 56
 DB 6 FVTQSSRMETLGLILFMASLSBAFNIDEPHLPFGTEDEDFRGSVYOTEGNRKQI 65
 QY 57 LVGAPWDPSGDRRGDVYRCVPGAHNAFCARGLGDIQGNSSHPANMHLGMSLLET 116
 DB 66 IVGAPLENSA---GEMVSC---ADLQSKR-----LGRPSESVRF--FGMSAAYS- 110
 QY 117 DGGFMACAPLMSRAAGSVSSGICARVDASFOPOGSLAPTAQRC--PTYMDVIVLVDGS 175
 DB 111 -SAALTSCSPYFAHECDGNSYINGVYQFNSLQAVSNFTAAVQECSSKREVNVLVFLFDGS 169

QY 622 DGGDLVAVAGAGAAIILSRPIVHLTPSLVETPOAIVQRCRRRGGAEVCLTAALC 681
 DB 647 SGGDLADIVTGGSDVAVLRSRPVOLLTYSMTFPLPALPAFMD-----KMDVELC 697
 QY 682 FOYTST---PGRMDHOFYMRFTASLDEWTAGARAFDGGORL-----SPRLRL 729
 DB 698 FKVDSSAVSEPEGL--RGMSLNFTVOVD-----VTXKORLOCADRSCCGSLMKW 746
 QY 730 SVGNVTCGOLHFNVLDT-----SDYLRPVALVTFTALDNTK-----PGVLNE--GSPRS 778
 DB 747 SGGSSICE--HFLITETBELCEDDCFSNTITIVSTFOTSEGRNHPNPIIDHYEPBA 804
 QY 779 IOKLVPFSDCCGDNCEVTDVLQVNMIDGRSKAPFVVGGRKYLVTLEENKEXAY 838
 DB 805 IFQL--PYEKDCCKNVKFCIAEIQLTATISQOD-----LVVGTIKVEVTNMINSLTNGSEDSY 857
 QY 839 NTSLSLIFSRNLHLASLTPQRESPIVECAAPSAHARL-----CSVGHVPFO--TGAKYT 891
 DB 858 MTNMAINYPRLQPKKI--QKPLSPDIQCDPKPVASVLMNCKIGHPIIKRGSVNVSYT 915
 QY 892 FILEFFSCSSLSLOVFGKLTAS-----SDSLERNGLQENTAO--TSAYIQY--BPHLLF 943
 DB 916 WQLE-----ESIFPNRTADITVTISNKSLSARETHSLQFPHAFIAYLSRPSVWY 966
 QY 944 ---SSESTLHRYEVHAPYGLTPVGPGPEFKTTLRVQNLGCVVSGLIISALLPAVAHGNY 1000
 DB 967 NMTSQSSSDHK-----EFPFNV-----HGENH 988
 QY 1001 FLSLSQVITNNASCIVQNLTEPPGPPVHPELO-----HTNRLNGSN--TQCVVACHIG 1053
 DB 989 FGAVFQL-----QICV-----PILRADLIQIRVKHLTKQATCTCQSQEPCTGSD 1034
 QY 1054 QAKGTGVSAGLRLVHNEFFRAKFKSLTVSTFELG-----TEGVSVLQLEASRWSE 1108
 DB 1035 PVOVNEHMSVIAITSNK-----ENVTAAEISMGTKOLLRISLQILGELISFNK 1087
 QY 1109 SLELV-----QTRPLI-----SLWLLIGSVLGLLLALLVFCIMKLGFF--ARK 1153
 DB 1088 SYEGNAENHRTKIVFLKEKPHSLPLIIGSSIGLLVLVIAILFKCGFFFRKVK 1147
 QY 1154 KIPBEKREBKE 1166
 DB 1148 QNLSESVRAQLK 1160

RESULT 14
 ID Q28984 PRELIMINARY; PRT; 920 AA.
 AC Q28984;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE CD11b (fragment).
 GN CD11B.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
 OX NCBI_TaxID=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RL Lee J.-R., Schook L.B., Rutherford M.S.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U40072; AAB16869.1; -
 DR HSSP; P11215; 1A8X.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF001839; FG-GAP; 4.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; vwa; 1.
 DR PROSITE; PS50234; VWF_A; 1.

FT NON TER 1 1
 FT NON TER 920 920
 SQ SEQUENCE 920 AA; 102440 MW; E96CC51E350DD5AC CRC64;
 Query Match 14.5%; Score 883.5; DB 6; Length 920;
 Best Local Similarity 29.1%; Pred. No. 6; 9e-58;
 Matches 288; Conservative 165; Mismatches 376; Indels 161; Gaps 39;
 QY 157 PTAQR--CPY--MDVIVIVLDGNSI--YPMSEVQTFRLRVGLTFDPQIQVGLVOYGS 212
 DB 5 PALAGCCPQOESDIFLIDGSSINRLOPQRKEKVESTVMGO--FOKSKTLFALQYSED 62
 QY 213 PVHEMSLDGR--TKEEVRAAKNLSRREGREYKTAQAINVACTEGFSQSHGRPEAR 269
 DB 63 FYTHFTFNDPKRNPSPKLLVRIPIQLGR---THTAGIRKRVNELPHSKGARENALK 118
 QY 270 LLVYVTDGSHDGEELPALKACEGR--VTRYGLAV--LGHYLRQRDRPSFLREIRT 324
 DB 119 ILVVTIDKEKF--GDELGYEDVLPEDARKGVIRYVGVDAFNSWKSRE-----ELNT 169
 QY 325 IASDDERFEFNVTDALVTDVLDGRIFGLEGSHAENESSFGLMSQIGFSTRLLKD 384
 DB 170 IASKRCGDHVPOVNNFEAVKTIQNLQEKTAIEGTQSTISFCEMSQEGFSALITSN 229
 QY 385 GILFGMVGAYDM--GGSVLMLEGGHRLPFRMALDEDEPPALQNHAAVLYGYSSMLRGC 443
 DB 230 GPLLGAVGSPDWAGAPFLHMPKDRVIFINTTRVDSM-----NDAYLGAYV--EVILRNQ 282
 QY 444 RRLFLSGAPRFHRKRVIAFOLKKGAVRVAOSLOGEQIGSYFSGELCPDTRDCTDV 503
 DB 283 AOSLVLAGAPRYOHTGLVWMFK--QNSGAMEKNADIKGSDIGSYFSGALSVDNRGSSDL 341
 QY 504 LIVAAPMFLG-----PQNKETGRVYVYLVGOOSLLTLGLOPEPPQDARF 551
 DB 342 VILGAPHYEQRRGQVSVCPYPOGRAKQCRVILCEQG-----HPW---SRGCA 389
 QY 552 AMGALPDINODGFADAVAGAPLEDGHOGALYLYHGT--OSGARPHBQRIIAASMPHALSY 610
 DB 390 ALTALGDVNGDKLTUVALGAPGQDNRGAVVLYFHGTSSELGISPSHQRIAGQSPLQY 449
 QY 611 FGRSVDGRLDLDGDLVAVAGAGAAIILSRPIVHLTPSLVETPOAIVQRCRRRG 670
 DB 450 FGQSLSGGODLTMDELMDLAVGAQGHVFLRSQPLRAEASVVFPRVARVNYCROOA 509
 QY 671 QEA--VCLTAALCFQVTSRTPGRMDH-----QFYVRFASLDEWTAGARAFDGGORLS 723
 DB 510 AKTOIAGEVOYQLQKST--WDLRREGDQDSITTYDLADPGRPHRAVPEEL--KN 564
 QY 724 PRLRLSVG--NVTCEQLHFNHLD--TSDYLRPVALVTFTALDNTKPG-----DVINEG 774
 DB 565 TRQOTQTLGSRKCEHLALWLPDCEVDSVTPVLRNLPSL--VGRPASSFGNLRPLVAVD 622
 QY 775 SPTSIQKLVPSKDCGPNCEVTDVLQ---VNMIDGRSKAPFVVGGRKYLVTLEENKEXAY 831
 DB 623 AGRLFTALFPEPKNGNSICQDDLSITFSFMSLDT-----LVVGGPRDLKTYLVR 674
 QY 832 NRKENAVYTSLSIISRMHLASLTPOR-----SSPIKECAAPSAHARLCSV 879
 DB 675 NOGEDSYRTQVTFPSPDLSYRKVSTSONQSQSRWRLACSDVTE--ESTALKSTSCSI 733
 QY 880 GHVPQTKAYTFLLEFFSCSSLSLOVFGKLTASDSLERNGLQENTAO-----930
 DB 734 NHPIDPNDSSEVTFVNPDPADFLGY--KLLKAVNTSENNMPSNKTEFOLELVPKY 790
 QY 931 -----TSAYIOYEHLLFSSESTLHRYEVHAPYGLTPVGPGPEFKTTLRVQNLGCVVSG 984
 DB 791 AVYVVVTSLEVTXKFNFTASEKTRHVE--HOY-----QFNNLG---QRK 831
 QY 985 LIISAL-----LPVAAGNVFLTSQVITNNASCIVQNLTEPPGPPVHP---BELQHT 1035
 DB 832 LPISVFWVPVRLNRYVWDOQVYTFSONLSRSC-----TEELGPR--HSDFLEKLOKT 884
 QY 1036 NRLNGSNTQOCVVRCHLQOLAKGTEVSGL 1065

```

Db      885 PVLNCSIAVCXKICQDIPSGIQEIBLXVLT 914

RESULT 15
Q9WUF8      PRELIMINARY; PRT; 895 AA.
ID Q9WUF8
AC Q9WUF8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Itgae protein (Fragment).
GN ITGAE.
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10095;
RN [1]
RP SEQUENCE FROM N.A.
RA Schoen M.P., Arya A., Murphy E.A., Adams C.M., Strauch U.G.,
RA Agace W.W., Marsal J., Donchue J.P., Her H., Beter D.R., Olson S.,
RA Lefrancois L., Brenner M.B., Grusby M.J., Parker C.M.;
RT "Mucosal T lymphocyte numbers are selectively reduced in integrin
RT alphae (CD103) deficient mice."
RL Submitted (Mar-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF133085; AAD30063.1; -.
DR EMBL; AF133070; AAD30063.1; JOINED.
DR EMBL; AF133071; AAD30063.1; JOINED.
DR EMBL; AF133072; AAD30063.1; JOINED.
DR EMBL; AF133073; AAD30063.1; JOINED.
DR EMBL; AF133074; AAD30063.1; JOINED.
DR EMBL; AF133075; AAD30063.1; JOINED.
DR EMBL; AF133076; AAD30063.1; JOINED.
DR EMBL; AF133077; AAD30063.1; JOINED.
DR EMBL; AF133078; AAD30063.1; JOINED.
DR EMBL; AF133079; AAD30063.1; JOINED.
DR EMBL; AF133080; AAD30063.1; JOINED.
DR EMBL; AF133081; AAD30063.1; JOINED.
DR EMBL; AF133082; AAD30063.1; JOINED.
DR EMBL; AF133083; AAD30063.1; JOINED.
DR EMBL; AF133084; AAD30063.1; JOINED.
DR HSP; P11255; IABX.
DR InterPro; IPR00413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 4.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR01185; INTEGRIN.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00234; VWF_A; 1.
FT NON_TER
FT NON_TER
SQ SEQUENCE      895 AA; 98266 MW; BEEBA14A754ADAIE CRC64;

Query Match      14.5%; Score 883; DB 11; Length 895;
Best Local Similarity 30.5%; Pred. No. 7,2e-58;
Matches 295; Conservative 154; Mismatches 333; Indels 186; Gaps 48;

QY      23 ENLDEHH---PRLFGPPAEAFEGYVLOHV--GGGQRMVLVGAPWDGPGSGRRGDVYRCP 77
DB      7 FNMDDVMAMVTLQCGAPAV---LSSLHLDPSSNNOTCLIVAR---RSSNRNTALYRCA 60
QY      78 VG-GAHNAPCAR-GHL-----GDYOLGNSHRAVNMHLGMSLLETDDGCFMACALMGR- 130
DB      61 ISISPEDEIACQPEVHEHCMPKGRYQ-----GVTIV--GNHNGVLVCIVQVQAK 105
QY      131 -----AC-----GSSVFSSGICARVDASFQPGSLA 156
DB      106 FRLNSLSELTGACSLTLPNLDLQAVFSDLEGFDPGAFHVDGDCYCRSKGS---TGEEK 162
QY      157 PTAQRCPTV-----MDVVIYLDGNSNITPMSEVQ-----TFLRLVGLKFLIDPE 200

```

```

Db      163 KSARRRRIVEEEDDEEDGTEIAIYLDGSGSLIEP--SDFOKANFISTMMRNFYEKCF----- 217
QY      201 QIOVGLVQYGESPVHEWGLGDFR-----TKEEVVRANKLSRRRGRETAKTAQAIMVACT 254
Db      218 ECFNALVQYGAIVQIEFPLQESRDINSLAKVQSIYQKEV-----TKTRAMQHVLVD 270
QY      255 EGFSGSHGRRPEARLVLVVTGDESH--DGEELPAALKACEAGRYTRYGIAVLGHYLROR 313
Db      271 NIFIPRSRGRKALKVMVVLTDGDFGDPPLNLTITVINSPPKQGVVRFAGV--GDAFKN-- 327
QY      314 DPSSFLREIRITIASPPDRFPFNVTDEALLDIYDALDRIPLGLEGSAENESSGLEMS 373
Db      328 --NMTYRELKLIASDPKKAHTFKYTNYSALDGLLSKQQRIVHMGIVGD--ALQYOLA 382
QY      374 QIGFSTHLKKG--IFGAWGAYDW--GSGVLM--LEGHRLFPFRMLLEDFPALQNHAY 430
Db      383 QTFPSAQILDGQVLLGVGAFPMNSGGLLSTQNGRCFLNQTKED---SRIVQYISY 438
QY      431 LGYSVSMILRGRRLLFSGAPRRHRGKVIAFOLKDGAVRVA--QSLQGEQISYFGS 488
Db      439 LGYSL-AVLHKAHGISYVAGAPRHKLRCAY--FELRKEDREDAFVRRIEGQMGSYFGS 495
QY      489 ELCPIDTRDGTDTVLVAAMPFLCPQKKEGRVYVYLVGOQ--SLITLQGTIQPEPP--QD 546
Db      496 VLPVDIDMDGTTDFLVAAPFY--HIRGEBGRVYVQVPPQDASFLAHLTSGHGLTN 553
QY      547 ARFGFAMGALDPLNDGFDADVAVGAPE-----DG-HOGALYLYHGTSQGVRRPRAORIA 600
Db      554 SRFGFMAAVGDINDKKTDTVAIGAPLEGFGAGDASGSIYITNGHSGGLYDSFSQOLR 613
QY      601 AASMPHALSYFGRSYDGRULDLDGDDLVDAVGAQGAAILSSRPVHLTPSLEVTPOAIS 660
Db      614 ASSVASGLHYGMSVSGGLDFNGDGLADITVGSRSAAVLNRPVVDLTIVSMTFTPDALP 673
QY      661 VV---QPRCKRRGCEAVCLTALCFQVTSRT---PGRWDHOFWRFTASLDEMTAGARA 713
Db      674 MVFIKMDVNV-----LCFEVDSSVVASEPGU--REMPUNFTVDVDVTKORQRL 719
QY      714 AFDGSGQRSLSPRLSLSVGNTVCEQLHFNVLDT-----SDYLRPVALTFTVFDLNT 764
Db      720 QCEDSSGQSCLR--KMGSGFLCE--HFWLISTELCEDEDFPSNITTIVTTEFQSGRR 776
QY      765 TKPGFVLNE--GSPTSIQLVPEFSKDCGPDNCCVTDLVQVNMIDIRGSRKAPFVARGRR 823
Db      777 DYPNPTLDHYKEPSAIFQL--PYEKDCKKVKVFCIAEIQLTIVISQGE-----LVVGATKE 829
QY      824 VLVSTTLNKKENAYNTSLSIIFSRNLHLASLTLPQRESPIK--VECAAPSAHARL-----C 877
Db      830 VTMNISLTNSGEDSDSYMIMALNYPRLIQFKKI---QKPVSPDVQCDPDKPVASVLVWNC 885
QY      878 SVGHVPFQ 885
Db      886 KIHFILK 893

```

Search completed: July 16, 2003, 07:55:23
Job time : 126.819 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 129.667 Seconds

(without alignments)
1199.256 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPVTHTLPLVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq 101002: *
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6106	100.0	1167	21	AAV32242
2	6040	98.9	1167	22	AAAB64584
3	6031	98.8	1152	22	AAAB64584
4	5965	97.7	1152	22	AAAB64584
5	5904.5	96.7	1132	21	AAV32243
6	2439.5	40.0	1188	22	AAU14231
7	2439.5	40.0	1188	22	AAAB50065
8	2439.5	40.0	1188	23	AAU10551
9	2438.5	39.9	1188	22	AAU14467
10	2429	39.8	1189	21	AAAB5582

11	2429	39.8	1189	22	ABG12949
12	2422	39.7	1188	22	AAAB50087
13	2422	39.7	1188	23	AAU10552
14	2403.5	39.4	1188	22	AAAB30929
15	2254	36.9	1034	21	AAAB25590
16	1894.5	31.0	1180	23	ABAB90788
17	1893	31.0	1179	23	ABAB90759
18	1831	30.0	1183	20	AAV07728
19	1730	28.3	1183	20	AAV07729
20	1727.5	28.3	707	22	AAU19663
21	1727.5	28.3	707	23	ABP47883
22	1674	27.4	1367	19	ABP70542
23	1476	24.2	303	21	AAV32242
24	1363.5	22.3	979	22	ABG29239
25	1131.5	18.5	1161	16	AAV78166
26	1131.5	18.5	1161	18	AAW23049
27	1131.5	18.5	1161	19	AAV78285
28	1131.5	18.5	1161	19	AAV65089
29	1131.5	18.5	1161	19	AAV57491
30	1131.5	18.5	1161	20	AAV73342
31	1131.5	18.5	1161	21	AAV07359
32	1131.5	18.5	1161	23	ABG61468
33	1126	18.4	1161	18	AAW23064
34	1126	18.4	1161	19	AAV72837
35	1126	18.4	1161	19	AAW65106
36	1126	18.4	1161	20	AAV73343
37	1126	18.4	1161	21	AAV07376
38	1126	18.4	1161	23	ABG61485
39	1124.5	18.4	1161	16	AAV78169
40	1124.5	18.4	1161	18	AAW23062
41	1124.5	18.4	1161	19	AAV72824
42	1124.5	18.4	1161	19	AAV60004
43	1124.5	18.4	1161	21	AAV07374
44	1124.5	18.4	1161	23	ABG61483
45	1118.5	18.3	1161	19	AAV65104

ALIGNMENTS

RESULT 1	AAV32242	standard; Protein; 1167 AA.
ID	AAV32242	standard; Protein; 1167 AA.
XX	AAV32242;	
AC	AAV32242;	
DT	15-FEB-2000	(first entry)
XX		
DE	Human integrin subunit alpha-10.	
XX		
KW	Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;	
KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis;	
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;	
KW	fibroblast; vaccine; marker.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT	Protein	/note= "signal peptide"
FT	Protein	23..1145
FT	Protein	/note= "mature protein"
FT	Domain	23..1120
FT	Domain	/note= "extracellular domain"
FT	Domain	1121..1145
FT	Domain	/note= "transmembrane domain"
FT	Domain	1122..1167
FT	Domain	/note= "cytoplasmic domain, specifically claimed in Claim 21"
FT	Domain	162..359
FT	Domain	/note= "I-domain"
FT	Binding-site	494..502
FT	Binding-site	/note= "cation binding site motif"

Novel human diapo
Murine A259. Mus
Murine A259. polype
Amino acid sequenc
Protein encoded by
Rat Tumour Endothe
Human Tumour Endot
Armenian hamster a
Armenian hamster a
Human novel extrac
Human polypeptide
Integrin alpha-2 c
Mouse integrin sub
Novel human diapo
Human beta-2 integ
Human beta-2 integ
Human alpha-d. Ho
Human Beta-Integri
Human beta2_2 integ
Human alpha_d_1. R
Rat alpha d polype
Rat alpha d protei
Rat Beta2_integri
Rat beta-integrin

FT Binding-site 558..566
 FT /note= "cation binding site motif"
 FT Binding-site 620..628
 FT /note= "cation binding site motif"
 FT Modified-site 98
 FT /note= "N-glycosylated"
 FT Modified-site 336
 FT /note= "N-glycosylated"
 FT Modified-site 364
 FT /note= "N-glycosylated"
 FT Modified-site 733
 FT /note= "N-glycosylated"
 FT Modified-site 839
 FT /note= "N-glycosylated"
 FT Modified-site 921
 FT /note= "N-glycosylated"
 FT Modified-site 1018
 FT /note= "N-glycosylated"
 FT Modified-site 1039
 FT /note= "N-glycosylated"
 FT Modified-site /note= "N-glycosylated"
 PN MO951639-A1.
 PD 14-OCT-1999.
 XX
 PF 31-MAR-1999; 99WO-SE00544.
 XX
 PR 02-APR-1998; 98SF-0001164.
 PR 28-JUN-1999; 99SE-0000319.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 PI Lundgren-Akerlund E;
 DR WPI; 2000-052639/04.
 DR N-PSDB; AA234719.
 PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation -
 XX
 XX Claim 1; Fig 6; 90pp; English.
 CC This sequence represents novel human chondrocyte integrin subunit
 CC alpha-10 (ISa10). A splice variant is given in AAY32243. The
 CC invention relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can
 CC be used as markers or targets of all types of cells, e.g. of
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:
 CC for treating pathological conditions involving ISa10, such as
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage,
 CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). ISa10 binding entities can be used to
 CC determine the differentiation-state of cells during embryonic
 CC development, angiogenesis or development of cancer, in pathological
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
 CC in tissue regeneration or in therapeutic and physiological repair
 CC of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. ISa10
 CC polynucleotides, vectors, host cells and methods of producing
 CC recombinant ISa10 are also claimed.
 XX
 SQ Sequence 1167 AA;

Query Match 100.0%; Score 6106; DB 21; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MELPVYTHLEPLVLTGLCSFNNIDHHPRLPFGPPPEAFGYSLVQHYGGGQRMVGA 60
 DB 1 MELPVYTHLEPLVLTGLCSFNNIDHHPRLPFGPPPEAFGYSLVQHYGGGQRMVGA 60
 QY 61 PMDGSGDRGDVYCPVGAHANAPCAKHGLGDYOLGNSSHPAVMHGLMSLLETDGCG 120
 DB 61 PMDGSGDRGDVYCPVGAHANAPCAKHGLGDYOLGNSSHPAVMHGLMSLLETDGCG 120
 QY 121 FMACAPLWSRACSSVFSSGICARVDASFOGSLAPTAORCPYMDVYIVLDGNSIYP 180
 DB 121 FMACAPLWSRACSSVFSSGICARVDASFOGSLAPTAORCPYMDVYIVLDGNSIYP 180
 QY 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVOYGESPVHEMSLGDRTKEEVYRAKNTSRPG 240
 DB 181 WSEVQTFRLRLVKGKLFIDPEQIQVGLVOYGESPVHEMSLGDRTKEEVYRAKNTSRPG 240
 QY 241 RETKTAQAIMVACTGEGFSQSHGGRPEAARLIVVYTDGSHDGEELPAALKACEAGRVPY 300
 DB 241 RETKTAQAIMVACTGEGFSQSHGGRPEAARLIVVYTDGSHDGEELPAALKACEAGRVPY 300
 QY 301 GIAVLGHYLRQRDPSSFLREIRITIASDPDERFPFNVTDEAALTDIVDAGDRIFGLEGS 360
 DB 301 GIAVLGHYLRQRDPSSFLREIRITIASDPDERFPFNVTDEAALTDIVDAGDRIFGLEGS 360
 QY 361 HAENSSFGLEMSQIGFSTHRLKDGILFGMVCAIYMGSSVLMLEGGHRLFPFRMALEDEF 420
 DB 361 HAENSSFGLEMSQIGFSTHRLKDGILFGMVCAIYMGSSVLMLEGGHRLFPFRMALEDEF 420
 QY 421 PPALONHAAVYIGYSVSMILRGGRFLFSGARFRRGKVIYFOUKKQCAVVAQSLQGE 480
 DB 421 PPALONHAAVYIGYSVSMILRGGRFLFSGARFRRGKVIYFOUKKQCAVVAQSLQGE 480
 QY 481 QIGSYFGSELCPLDTRDQDGTDLVLAAPMFLGPONKETGRVYVYLVOGQSLTLTQGLQ 540
 DB 481 QIGSYFGSELCPLDTRDQDGTDLVLAAPMFLGPONKETGRVYVYLVOGQSLTLTQGLQ 540
 QY 541 PEPPDARFGFAMGALPDLDNDFADVAAGALEBHGQALYLYHGTOSGVRRPRAQRIA 600
 DB 541 PEPPDARFGFAMGALPDLDNDFADVAAGALEBHGQALYLYHGTOSGVRRPRAQRIA 600
 QY 601 AASMPHALSYFRRSDVRGLDLDGDDLVAVAGAGAAIILSSRPVHLTPSLEVPQAI 660
 DB 601 AASMPHALSYFRRSDVRGLDLDGDDLVAVAGAGAAIILSSRPVHLTPSLEVPQAI 660
 QY 661 VVQRDCRRRGQAVCLTAALCFQVTSRTTPGRWDQFYMRFTASLDEWTAGARAAFDGSGQ 720
 DB 661 VVQRDCRRRGQAVCLTAALCFQVTSRTTPGRWDQFYMRFTASLDEWTAGARAAFDGSGQ 720
 QY 721 RLSPPRLRLSVGNVTCEQHLHFVVDTSYLRLVALTYVPALDNTTKRGPVINEGSPSTIQ 780
 DB 721 RLSPPRLRLSVGNVTCEQHLHFVVDTSYLRLVALTYVPALDNTTKRGPVINEGSPSTIQ 780
 QY 781 KLVPSKDCGPPNECVTDVLQVNMDDIRSRAPPVVGGRRKVIYSTLLENRKNAYNT 840
 DB 781 KLVPSKDCGPPNECVTDVLQVNMDDIRSRAPPVVGGRRKVIYSTLLENRKNAYNT 840
 QY 841 SLSIIFSRNLHSLTTPQRESPIKVECAPSAHARLCSGVHVPOTGAKVTFLEFEFSC 900
 DB 841 SLSIIFSRNLHSLTTPQRESPIKVECAPSAHARLCSGVHVPOTGAKVTFLEFEFSC 900
 QY 901 SLSLVQVRGKLTASSDSLEKNTLOENRQTSAYIQYEPHLLFSEESTLHREVPYGLT 960
 DB 901 SLSLVQVRGKLTASSDSLEKNTLOENRQTSAYIQYEPHLLFSEESTLHREVPYGLT 960
 QY 961 PVGPPEPFTTLRYVNLGCVVYVSGLIISALLPAVHAGNYFLSLQVTTNNASCIYONLT 1020
 DB 961 PVGPPEPFTTLRYVNLGCVVYVSGLIISALLPAVHAGNYFLSLQVTTNNASCIYONLT 1020
 QY 1021 BPPGPPVHPEELOHTNRUNGSNTOCVVRCHLGQLAKGTEVSVGLRLVHNEFFRARKK 1080

Query Match	Best Local Similarity	98.9%	Score 6040	DB 22	Length 1167
Matches 1155	Conservative	3	Mismatches	9	Indels 0; Gaps 0

QY	1	MELFPVTHLFLPLVFLTGLCSPPFNULDHNRPLFGPPPEAEFGYSVLQHVGGGQKMLVGA	60
Db	1	VELFEPVTHLFLPLVFLTGLCSPPFNULDHNRPLFGPPPEAEFGYSVLQHVGGGQKMLVGA	60
QY	61	PMDEPSGDRDQVRCVPVGAHNP.CAKGHLGDTQLGNSSHPANMHLGMSLLETDGGG	120
Db	61	PMDEPSGDRDQVRCVPVGAHNP.CAKGHLGDTQLGNSSHPANMHLGMSLLETDGGG	120
QY	121	FMACAPLMSRACGSSVFSSGICANVDASFQVQSLAPTAQCPITYMDVIVLDSNSIYP	180
Db	121	FMACAPLMSRACGSSVFSSGICANVDASFQVQSLAPTAQCPITYMDVIVLDSNSIYP	180
QY	181	MSEVQTEFLRRLVGKFLTDPEDQIQLGYQYGSPPVHEMSLDPFRKKEEYVRAKULSREG	240
Db	181	MSEVQTEFLRRLVGKFLTDPEDQIQLGYQYGSPPVHEMSLDPFRKKEEYVRAKULSREG	240
QY	241	RETTAQAQIMACIEGFSQSHGSPPEARLLVVTDSHSGEELPALKXCEGRVTRY	300
Db	241	RETTAQAQIMACIEGFSQSHGSPPEARLLVVTDSHSGEELPALKXCEGRVTRY	300
QY	301	GIAVLGYHLRQRPDSSFLREIRTIASDPDERPFENVTDEALATDIDALGDRI.FGLEGS	360
Db	301	GIAVLGYHLRQRPDSSFLREIRTIASDPDERPFENVTDEALATDIDALGDRI.FGLEGS	360
QY	361	HANESSFGLEMSQIGFSTHRLKXGILFGVAVADWGSVLMEGHRLEPPRALDEDF	420
Db	361	HANESSFGLEMSQIGFSTHRLKXGILFGVAVADWGSVLMEGHRLEPPRALDEDF	420
QY	421	PPALQNHAAVLYGYSSMLLRGGRRLPLSGAPPRRHNGKVIAFOLKXQDQAVRAVQSLQGE	480
Db	421	PPALQNHAAVLYGYSSMLLRGGRRLPLSGAPPRRHNGKVIAFOLKXQDQAVRAVQSLQGE	480
QY	481	QIGSYFSGELCPLTDRDQTTDVLVAAPMFLGQONKETGAVVYVLVGOQSLTLTQGLQ	540
Db	481	QIGSYFSGELCPLTDRDQTTDVLVAAPMFLGQONKETGAVVYVLVGOQSLTLTQGLQ	540
QY	541	PEPQODARFGFAMGALPDLDQDGFADVAVGA.PLEDHQGALYLVHGTQSVRPHPAQIA	600
Db	541	PEPQODARFGFAMGALPDLDQDGFADVAVGA.PLEDHQGALYLVHGTQSVRPHPAQIA	600
QY	601	AASPHALSTYGRSVDRGLDIDGDDLYDVAVGAQGAALLSSRPIVHLTBSLEVTPOAIS	660
Db	601	AASPHALSTYGRSVDRGLDIDGDDLYDVAVGAQGAALLSSRPIVHLTBSLEVTPOAIS	660
QY	661	VVQDRCRRGGEAVALC.FQVTSRTPCGMWHQFVFRFTASLIDEMWAGARAA.FDGSQ	720
Db	661	VVQDRCRRGGEAVALC.FQVTSRTPCGMWHQFVFRFTASLIDEMWAGARAA.FDGSQ	720
QY	721	RLSRRLRLSLSGANTCEOLHFHVLDTDYLR.PVALTYT.FALDNTTKBPVVLNEGSPISIQ	780
Db	721	RLSRRLRLSLSGANTCEOLHFHVLDTDYLR.PVALTYT.FALDNTTKBPVVLNEGSPISIQ	780
QY	781	KLVPSSKDCGPDNECVTDVLQVNMDIRGSRKAPVYVGRRKVLVSTTLENRKENAYNT	840
Db	781	KLVPSSKDCGPDNECVTDVLQVNMDIRGSRKAPVYVGRRKVLVSTTLENRKENAYNT	840
QY	841	SLSTIIFSRNLHLASLTQRESPIVECECAPAHARLCSVGHPIVQTGA.KVFLLEFFESC	900
Db	841	SLSTIIFSRNLHLASLTQRESPIVECECAPAHARLCSVGHPIVQTGA.KVFLLEFFESC	900
QY	901	SSLISQVFGKLTASDSLERNGTLOEONTAOTSAIYOVEPHLL.FSSESTLHRYEVPYTL	960
Db	901	SSLISQVFGKLTASDSLERNGTLOEONTAOTSAIYOVEPHLL.FSSESTLHRYEVPYTL	960
QY	961	PVGGEPEFKTTLAVQNLGCVVSGLLISALLPAVAHGNVFLSLSOVITNNASCIVQNL	1020
Db	961	PVGGEPEFKTTLAVQNLGCVVSGLLISALLPAVAHGNVFLSLSOVITNNASCIVQNL	1020
QY	1021	EPPEPVPHPRELOHTNRLNGSNTQCVVRCGLQGLAGTEVSUGVGLLUVHNEFPRAAKFK	1080
Db	1021	EPPEPVPHPRELOHTNRLNGSNTQCVVRCGLQGLAGTEVSUGVGLLUVHNEFPRAAKFK	1080

QY 1081 SLTVVSTFELGTEGSSVLQTEASRWSESLLEVQTRPILISIMILIGSVLGGILLALL 1140
 DB 1081 SLTVVSTFELGTEGSSVLQTEASRWSESLLEVQTRPILISIMILIGSVLGGILLALL 1140
 QY 1141 VFCLMKLGFPFAKKKIPDEEKREKLEQ 1167
 DB 1141 VFCLMKLGFPFAKKKIPDEEKREKLEQ 1167
 RESULT 3
 AAB64657
 ID AAB64657 standard; Protein; 1152 AA.
 XX AAB64657;
 AC 22-MAR-2001 (first entry)
 DT Human secreted protein BLAST search protein SEQ ID NO: 167.
 DE
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antiatherogenic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 PN MO200077197-A1.
 PD 21-DEC-2000.
 PF 01-JUN-2000; 2000MO-US14934.
 PR 11-JUN-1999; 99US-0138599.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 DR WPI; 2001-032312/04.
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PR used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; Page 543-546; 558bp; English.
 XX
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is a search
 CC result from a BLASTX homology search. The genes and proteins are useful
 CC for preventing, ameliorating or treating medical conditions, e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 XX
 SQ Sequence 1152 AA;
 Query Match 98.8%; Score 6031; DB 22; length 1152;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MELPVTYHLPPLVFLTGLCSPPNLDHHRPLFPGPPEAFEGYSVLQHWGGQRMVLVGA 60
 DB 1 MELPVTYHLPPLVFLTGLCSPPNLDHHRPLFPGPPEAFEGYSVLQHWGGQRMVLVGA 60

DB 1 MELPVTYHLPPLVFLTGLCSPPNLDHHRPLFPGPPEAFEGYSVLQHWGGQRMVLVGA 60
 QY 61 PNDGSGDRRGDYVRCVPGAHNAPCAKHLDYOLGNSSHPAVMHIGMSLLETGDCG 120
 DB 61 PNDGSGDRRGDYVRCVPGAHNAPCAKHLDYOLGNSSHPAVMHIGMSLLETGDCG 120
 QY 121 FMACAPLWSRACSSVFSSGICARVDASFOPGSLAFTAORCPTVMVYIVLDGNSIYP 180
 DB 121 FMACAPLWSRACSSVFSSGICARVDASFOPGSLAFTAORCPTVMVYIVLDGNSIYP 180
 QY 181 MSEVQTFRLRYGKLFIDPEQIQVGLVOYGSFVHENSIGDRTKEEVYRAKNLSRREG 240
 DB 181 MSEVQTFRLRYGKLFIDPEQIQVGLVOYGSFVHENSIGDRTKEEVYRAKNLSRREG 240
 QY 241 RETKTAQAIMVACTGFSOSHGRPEARLLVVVTGDESHDEELPALKACEAGRVTRY 300
 DB 241 RETKTAQAIMVACTGFSOSHGRPEARLLVVVTGDESHDEELPALKACEAGRVTRY 300
 QY 301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFENVTDAAITDIDVALGDRIFGLEGS 360
 DB 301 GIAVLGHYLRQRDSSFLREIRTIASDPDERFFENVTDAAITDIDVALGDRIFGLEGS 360
 QY 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMGAAYDWGGSVLMLEGGHRLFPERMALDEBF 420
 DB 361 HAENESSFGLEMSQIGFSTHRLKDGILFGMGAAYDWGGSVLMLEGGHRLFPERMALDEBF 420
 QY 421 PALQNHAAVLYGYSVSMILRGRLFTSGAPFRHGRKVIATFOLKXQDAVVAOSLOGE 480
 DB 421 PALQNHAAVLYGYSVSMILRGRLFTSGAPFRHGRKVIATFOLKXQDAVVAOSLOGE 480
 QY 481 QIGSYFGSELCPDIDRDGTTDVLVAAPMFLGPONKTEGRVYVYLVGQOSLTLTQGTLO 540
 DB 481 QIGSYFGSELCPDIDRDGTTDVLVAAPMFLGPONKTEGRVYVYLVGQOSLTLTQGTLO 540
 QY 541 PEPPDARFGFAMGALLPDLNODGFADVAVGAPLEHGHQALYLHGTOSQVAPHPHQAQIA 600
 DB 541 PEPPDARFGFAMGALLPDLNODGFADVAVGAPLEHGHQALYLHGTOSQVAPHPHQAQIA 600
 QY 541 PEPPDARFGFAMGALLPDLNODGFADVAVGAPLEHGHQALYLHGTOSQVAPHPHQAQIA 600
 DB 541 PEPPDARFGFAMGALLPDLNODGFADVAVGAPLEHGHQALYLHGTOSQVAPHPHQAQIA 600
 QY 601 AASMPHALSYFGRSVDGRDLDDGDLVDVAVAGAAIILSSRPVHLTPSLLEVTPAIS 660
 DB 601 AASMPHALSYFGRSVDGRDLDDGDLVDVAVAGAAIILSSRPVHLTPSLLEVTPAIS 660
 QY 661 VVORDCRRRGQAVCLTAALCFQVTSRTPGRWDFYMRFTASLDEMTAGARAAFDGSGQ 720
 DB 661 VVORDCRRRGQAVCLTAALCFQVTSRTPGRWDFYMRFTASLDEMTAGARAAFDGSGQ 720
 QY 721 RLSPRRRLASVANTCEQLHFRVLDTSYLREVALTVFALDNTTKPGPVINEGSPSTIQ 780
 DB 721 RLSPRRRLASVANTCEQLHFRVLDTSYLREVALTVFALDNTTKPGPVINEGSPSTIQ 780
 QY 781 KLVPSSKDCGPNCEVTDLVQVMDIRGSRKAPVVRGGRKVLVSTLENRKENAYNT 840
 DB 781 KLVPSSKDCGPNCEVTDLVQVMDIRGSRKAPVVRGGRKVLVSTLENRKENAYNT 840
 QY 841 SUSIIFSRNLHLASTLPQRESPIKVECAAPSAAHRLCSVGHVFPQTGAKVTFLEFEPSC 900
 DB 841 SUSIIFSRNLHLASTLPQRESPIKVECAAPSAAHRLCSVGHVFPQTGAKVTFLEFEPSC 900
 QY 901 SSLLSQVKGKLTASDSIERNCTLOENTAOQSAYIQYEPHLLFSESTLHREVPYCTL 960
 DB 901 SSLLSQVKGKLTASDSIERNCTLOENTAOQSAYIQYEPHLLFSESTLHREVPYCTL 960
 QY 961 PVGPPEPFKTLTVONLGCYVVGSLIISALPVAHAGNYFSLSQVITNNASCIYONLT 1020
 DB 961 PVGPPEPFKTLTVONLGCYVVGSLIISALPVAHAGNYFSLSQVITNNASCIYONLT 1020
 QY 1021 EPPGPVHPEELQHTNRLNGSNTOQVVRCHLGQIAKTEVSVGLRLVHNEFFPRAXFK 1080
 DB 1021 EPPGPVHPEELQHTNRLNGSNTOQVVRCHLGQIAKTEVSVGLRLVHNEFFPRAXFK 1080
 QY 1081 SLTVVSTFELGTEGSSVLQTEASRWSESLLEVQTRPILISIMILIGSVLGGILLALL 1140
 DB 1081 SLTVVSTFELGTEGSSVLQTEASRWSESLLEVQTRPILISIMILIGSVLGGILLALL 1140

QY 1141 VFCLWKLGFPAH 1152
DB 1141 VFCLWKLGFPAH 1152

RESULT 4
AAB64658
ID AAB64658 standard; Protein; 1152 AA.
AC AAB64658;
DE 22-MAR-2001 (first entry)
XX Human secreted protein BLAST search protein SEQ ID NO: 168.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
OS
XX WO200077197-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US14934.
XX
XX 11-JUN-1999; 99US-0138599.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE-) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komateoulis GA;
XX
XX WPI; 2001-032312/04.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX Disclosure; Page 547-551; 558pp; English.
XX
XX The invention relates to the isolation of genes AAF32757-F32803 encoding
XX the human secreted proteins AAB64549-B64594. The sequence is used as a
XX query sequence for doing BLASTX searches to identify homologous
XX sequences. The genes and proteins are useful for preventing,
XX ameliorating or treating medical conditions, e.g. by protein or gene
XX therapy. The genes are isolated from a range of human tissues disclosed
XX in the specification. The nucleic acids, proteins, antibodies and
XX (ant)agonists are useful in the diagnosis, treatment and prevention of:
XX (a) cancer; e.g. breast and ovarian cancer, and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
XX disorders such as myocardial ischaemias; (d) wound healing; (e)
XX neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections.
XX
XX Sequence 1152 AA;
SQ

Query Match 97.7%; Score 5965; DB 22; Length 1152;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1140; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MELPVTHTLFLPLVFTGLCSPPNLDHHPRLFPGPPEAFPGYSVLQHGQGMWLVGA 60
DB 1 VELPVTHTLFLPLVFTGLCSPPNLDHHPRLFPGPPEAFPGYSVLQHGQGMWLVGA 60

QY 61 PWDGSGDRRDVTRCPVGAHNAAPCAKGLADYIOLGNSSHPAVMHLGMSLLETDGCG 120
DB 61 PWDGSGDRRDVTRCPVGAHNAAPCAKGLADYIOLGNSSHPAVMHLGMSLLETDGCG 120
QY 121 FMACAPLMSRACGSSVFSSGICARVDASFQPGSLAPPAORCPYMDVIVLDGNSLYP 180
DB 121 FMACAPLMSRACGSSVFSSGICARVDASFQPGSLAPPAORCPYMDVIVLDGNSLYP 180
QY 181 MSEVOTFLRLVGLKFLIDPEQIQVGLVOYGSPVHEMSLGFRTKEEVRAAKNLSRREG 240
DB 181 MSEVOTFLRLVGLKFLIDPEQIQVGLVOYGSPVHEMSLGFRTKEEVRAAKNLSRREG 240
QY 241 RETKTAQAIMVACTEGFSQSHGPREARLLVVTVDGSHDGEELPAALKACEAGVTRY 300
DB 241 RETKTAQAIMVACTEGFSQSHGPREARLLVVTVDGSHDGEELPAALKACEAGVTRY 300
QY 301 GIAVGHILRRQRPSSFLREIRITASPDPRFPFNVTDEALTDIVALGDRITGLEGS 360
DB 301 GIAVGHILRRQRPSSFLREIRITASPDPRFPFNVTDEALTDIVALGDRITGLEGS 360
QY 361 HAENESSFGLMSQIGFSTHRLKDCIILFGMVGAIDMGSVLMBGHRLLFPRMALDEBF 420
DB 361 HAENESSFGLMSQIGFSTHRLKDCIILFGMVGAIDMGSVLMBGHRLLFPRMALDEBF 420
QY 421 PPALQNHAAVLYGYSVSMILRGRRFLSGAPRRHRKGVIAFOLKXDAVRAVSLQGE 480
DB 421 PPALQNHAAVLYGYSVSMILRGRRFLSGAPRRHRKGVIAFOLKXDAVRAVSLQGE 480
QY 481 QIGSYFSELCPDIDRPGTTDVLVAAAPMLGPNKKTGVVYVLYVQOQSLLTIOGLQ 540
DB 481 QIGSYFSELCPDIDRPGTTDVLVAAAPMLGPNKKTGVVYVLYVQOQSLLTIOGLQ 540
QY 541 PEPPODAFPGAMGALPDLDGDDLVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 600
DB 541 PEPPODAFPGAMGALPDLDGDDLVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 600
QY 601 AASMPHALSYGRSVDGRLLDGDLDVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
DB 601 AASMPHALSYGRSVDGRLLDGDLDVDAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
QY 661 VVORPCRRRGOEAVCLTALCFVTSRTPGMDHOFYMRFTASIDEMWAGARAAADGSGQ 720
DB 661 VVORPCRRRGOEAVCLTALCFVTSRTPGMDHOFYMRFTASIDEMWAGARAAADGSGQ 720
QY 721 RLSPPRLSLSGVNTCEQLHFHVLDTSDYLRLPALVTFALDNTTKPGPVINEGSPISIQ 780
DB 721 RLSPPRLSLSGVNTCEQLHFHVLDTSDYLRLPALVTFALDNTTKPGPVINEGSPISIQ 780
QY 781 KLVPSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKYLVTTLLENKENAYNT 840
DB 781 KLVPSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKYLVTTLLENKENAYNT 840
QY 841 SLSTIFSNLHLASITPQRESEPIKECAAPSAHARLCSVGHVPQTGAKVTLLEFEFSC 900
DB 841 SLSTIFSNLHLASITPQRESEPIKECAAPSAHARLCSVGHVPQTGAKVTLLEFEFSC 900
QY 901 SLSLSQVQKLTASSDSIERNGTLOENAOISAVIYOEPHLLFSESESTLHYEVHPTYL 960
DB 901 SLSLSQVQKLTASSDSIERNGTLOENAOISAVIYOEPHLLFSESESTLHYEVHPTYL 960
QY 961 PVGPGPEFKTLRLVONLCYVVSGLIISALLPAVAGHNYFLISQVITNNASCIVQULT 1020
DB 961 PVGPGPEFKTLRLVONLCYVVSGLIISALLPAVAGHNYFLISQVITNNASCIVQULT 1020
QY 1021 EPPGPVPAPELQHTNRLNGSNTQCOVVRCHLGOLAKTEVYSVGLRLVHNEFFRRAFX 1080
DB 1021 EPPGPVPAPELQHTNRLNGSNTQCOVVRCHLGOLAKTEVYSVGLRLVHNEFFRRAFX 1080
QY 1081 SLTVVSTPELGTGERSVQLTEASRWSSESLLEVQTRPILSLMILIGSVIGGLLIALL 1140
DB 1081 SLTVVSTPELGTGERSVQLTEASRWSSESLLEVQTRPILSLMILIGSVIGGLLIALL 1140
QY 1141 VFCLWKLGFPAH 1152

Db 1141 VFCLMKLGFFAH 1152

RESULT 5
AAV32243

ID AAV32243 standard; Protein; 1132 AA.

AC AAV32243;

DT 15-FEB-2000 (first entry)

DE Human integrin subunit alpha-10 splice variant.

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;

KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;

KW inflammation; therapy; cartilage; chondrocyte; osteoblast;

XX fibroblast; vaccine; marker; splice variant.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..22 /note="signal peptide"

FT Protein 23..1132 /note="mature protein"

PN MO9951639-A1.

PD 14-OCT-1999.

XX 31-MAR-1999; 99WO-SE00544.

PR 02-APR-1998; 98SF-0001164.

XX 28-JUN-1999; 99SE-0000319.

PA (ACT1-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

PI MPI; 2000-052639/04.

DR N-PSDB; AA234720.

XX New isolated integrin subunit alpha-10, used as a marker or target

PS molecule for cells during development, regeneration and pathological

XX conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

PT inflammation

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human

CC chondrocyte integrin subunit alpha-10 (ISa10). It is identical to

CC ISa10 (see AAV32242) except for deletion of amino acids 975-986. The

CC invention relates to a recombinant or isolated integrin heterodimer

CC (consisting of alpha-10 subunit in association with subunit beta

CC (consisting of beta-1). The heterodimer, subunit alpha-10 or splice

CC variant can be used as a marker or target of all types of cells, e.g.

CC of chondrocytes, osteoblasts and fibroblasts. They can also be used:

CC for treating pathological conditions involving ISa10, such as

CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;

CC for detecting the formation of cartilage during embryonal

CC development, physiological or therapeutic repair of cartilage,

CC or detecting regeneration of cartilage or chondrocytes during

CC transplantation of cartilage or chondrocytes; for selection and

CC analysis or for sorting, isolating or purification of chondrocytes

CC and for in vitro studies of differentiation of chondrocytes; and as

CC a target for anti-adhesive drugs or molecules in tendon, ligament,

CC skeletal muscle or other tissues where adhesion impairs the function

CC of the tissue (all claimed). ISa10 binding entities can be used to

CC determine the differentiation-state of cells during embryonic

CC development, angiogenesis or development of cancer, in pathological

CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,

CC in tissue regeneration or in therapeutic and physiological repair

CC of cartilage (claimed). A vaccine comprising the integrin

CC heterodimer or subunit alpha-10 is also claimed. ISa10

CC polynucleotides, vectors, host cells and methods of producing

CC recombinant ISa10 are also claimed.

XX Sequence 1132 AA;

XX Query Match 96.7%; Score 5904.5; DB 21; Length 1132;

XX Best local similarity 97.0%; Pred. No. 0;

XX Matches 1132; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

XX

QY 1 MELPVTHLFLPLVLTGLCSPPNIDENHPRLPFPPPEAEFGSYLVQVGGGQRMVLYGA 60

DB 1 MELPVTHLFLPLVLTGLCSPPNIDENHPRLPFPPPEAEFGSYLVQVGGGQRMVLYGA 60

QY 61 PWDGSGDRGGVYVCPVGAHNAACAGHIGDYVLGNSHPVAMNHGMSLLETDGGG 120

DB 61 PWDGSGDRGGVYVCPVGAHNAACAGHIGDYVLGNSHPVAMNHGMSLLETDGGG 120

QY 121 FMACAPLMSRACGSSVFSSGICARVDASFOPGSLAPPAORCPTYMDVIVLDGSSNSIYP 180

DB 121 FMACAPLMSRACGSSVFSSGICARVDASFOPGSLAPPAORCPTYMDVIVLDGSSNSIYP 180

QY 181 WSEVOTFLRLVYKLFIDPEQIQVGLVOYGSSPVHMSLGDPRYKEEVVRAKNSRREG 240

DB 181 WSEVOTFLRLVYKLFIDPEQIQVGLVOYGSSPVHMSLGDPRYKEEVVRAKNSRREG 240

QY 241 RETKTAQAIMVACTGFSQSHGSGREARLLVVTVDGSHDGEELPALKKCEAGRVRY 300

DB 241 RETKTAQAIMVACTGFSQSHGSGREARLLVVTVDGSHDGEELPALKKCEAGRVRY 300

QY 301 GIAVLGHLRLRRDSSFLREIRTIASDPDERFFPNVDEALTDIVDALGRIFGLEGS 360

DB 301 GIAVLGHLRLRRDSSFLREIRTIASDPDERFFPNVDEALTDIVDALGRIFGLEGS 360

QY 361 HAENSSFGLEMSQIGFSTHRLKQGLFGMGAYVWGSSVLMLEGGHRLFPFPRALDEBF 420

DB 361 HAENSSFGLEMSQIGFSTHRLKQGLFGMGAYVWGSSVLMLEGGHRLFPFPRALDEBF 420

QY 421 PPALONHAAYIGYSVSMWLRGRRFLPSGA PRPHRGKVIAPOLKKQGAARVAOSLGE 480

DB 421 PPALONHAAYIGYSVSMWLRGRRFLPSGA PRPHRGKVIAPOLKKQGAARVAOSLGE 480

QY 481 QIGSYFGESELCPLDTRDGTDDVLLVAA PMFLGPONKETRYYVVLVGOQSLLTLQGTIQ 540

DB 481 QIGSYFGESELCPLDTRDGTDDVLLVAA PMFLGPONKETRYYVVLVGOQSLLTLQGTIQ 540

QY 541 PEPPQDAFFGFAMGALPDLDGDFADVAVGAPLEBGHOGALVLYHGTSQVRRPRAORIA 600

DB 541 PEPPQDAFFGFAMGALPDLDGDFADVAVGAPLEBGHOGALVLYHGTSQVRRPRAORIA 600

QY 601 AASMHALSYFGRSDVGRDLDDGDLVAVAGAGAAILLSRPTVHLTPSLEVTPORIS 660

DB 601 AASMHALSYFGRSDVGRDLDDGDLVAVAGAGAAILLSRPTVHLTPSLEVTPORIS 660

QY 661 VVQRCRRRGGAEVCLTALCFQVTSRTTPGMDHOFYVFTASLDEWTAGARAADGSGQ 720

DB 661 VVQRCRRRGGAEVCLTALCFQVTSRTTPGMDHOFYVFTASLDEWTAGARAADGSGQ 720

QY 721 RLSPPRLRLSVGNVTCBOLHFRVLDTSYLRLPVALITVPALDNTTKPGPVINEGSPISIQ 780

DB 721 RLSPPRLRLSVGNVTCBOLHFRVLDTSYLRLPVALITVPALDNTTKPGPVINEGSPISIQ 780

QY 781 KLVPSSKCGPNNECVTDLVQVNMDDIGSKRAPVVGGRKIVSTLLENKKNANVT 840

DB 781 KLVPSSKCGPNNECVTDLVQVNMDDIGSKRAPVVGGRKIVSTLLENKKNANVT 840

QY 841 SLISIFSNLHLASLTPRESPIKVECAAPAHARLCSVGHVPOTGAKVTFLEFEFSC 900

DB 841 SLISIFSNLHLASLTPRESPIKVECAAPAHARLCSVGHVPOTGAKVTFLEFEFSC 900

QY 901 SLLSIVFGKLTASSDSLERNGTQENTAGTSAYIOYEPHLLFSSSELTARVHPYGTLL 960

DB 901 SLLSIVFGKLTASSDSLERNGTQENTAGTSAYIOYEPHLLFSSSELTARVHPYGTLL 960

Db 901 SSLSSQVFGKLTASSSLERNGLTQNTAAYIOYEPHLLFSSSTLHRYEHPYGL 960
 Qy 961 PVGPGPEFKTLRVQNLGCVVSGLLISALLPAVHAGNVPFLISQVITMNASCIYQNT 1020
 Db 961 PVGPGPEFKTLR-----TNNASCIYQNT 985
 Qy 1021 EPPGPPVHPEELQHTNRLNGSNTQCCVVRCHLGQAKGTEVSGILRLVHNEPFRRAK 1080
 Db 986 EPPGPPVHPEELQHTNRLNGSNTQCCVVRCHLGQAKGTEVSGILRLVHNEPFRRAK 1045
 Qy 1081 SLTVSTPELSTEEGSLVQLTASRWSSESLLEVQTRPILISLWILIGSLGLLALL 1140
 Db 1046 SLTVSTPELSTEEGSLVQLTASRWSSESLLEVQTRPILISLWILIGSLGLLALL 1105
 Qy 1141 VFCLWKLGFPAHKKIPEEKREKLEQ 1167
 Db 1106 VFCLWKLGFPAHKKIPEEKREKLEQ 1132
 RESULT 6
 AAU14231
 ID AAU14231 standard; Protein; 1188 AA.
 AC AAU14231;
 DT 24-OCT-2001 (first entry)
 DE Human novel protein #102.
 KM Human; novel protein. Antianemic; osteopathic; antiinflammatory;
 immunomodulatory; cytosstatic; neuroprotective; vulnery; nocotropic;
 anticonvulsant; antirheumatic; cerebroprotective; antifungal; antiviral;
 antibacterial; antiallergic; dermatological; haemostatic; antiaethmatic;
 thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 tissue regeneration; immune disorder.
 OS Homo sapiens.
 PN WO200155437-A2.
 PD 02-AUG-2001.
 PF 25-JAN-2001; 2001WO-US02623.
 PR 25-JAN-2000; 2000US-0491404.
 PA (HISE-) HISEQ INC.
 PI Tang YT, Liu C, Drmanac RT;
 DR WPI: 2001-451939/48.
 DR N-PSDB: AAS22536.
 PT Isolated peptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX
 PS Example 4; Page 578-581; 894pp; English.
 CC The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicite an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.

CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 SQ Sequence 1188 AA;
 Query Match 40.0%; Score 2439.5; DB 22; Length 1188;
 Best Local Similarity 43.2%; Pred. No. 7.5e-208;
 Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;
 Qy 1 MELPVTLLPLPLVEFLTGLCSFNLDEHHPRLPGPPAEFGYSVLQHVGGQRMVLVGA 60
 Db 1 MDLPRLVAVAMALSLMPGFTDTPNMDTRKPRVIGSRTAFQYTVQQHDISGNKMLVYGA 60
 Qy 61 PWDGSGRGDGYRCPIVGAHANAPCAAGHGLGVOLGSSHPAVMMHGMSLLETDGCG 120
 Db 61 PLETNQYQKTGDVYKCPV---IHGCTKLNLGRVTLNVSVSRKDMRGLSLATNPKNDS 117
 Qy 121 FMACAPLMSRACGSSVFSGGICARVDASFQOGSLAPPAOCPTVMDVYIVLDGSSNTYP 180
 Db 118 FLACSPPLMSHCGSSYTTTGCMCRVNSNFRSKVTAPALQRCQYMDVYIVLDGSSNTYP 177
 Qy 181 WSEVOTFLRLVGLKFLIDPEOIQVGLVOYGSVPVHWSLGDFTKEEVVRAKKNLSREG 240
 Db 178 WVEVQHFLINILKKFYIPGQIQVGVQYGDVHVEPLNDYRSVKDVVAASHIEQGG 237
 Qy 241 RETKTAQAIMACTGFSQSHGGRPEARLLVYVTDGSSHGEEPLPAALKACEARVTRY 300
 Db 238 TETRTAFGIIEPARSEAFQK--GGRGAKKVMIVITDGSHSDPDLKVIQOESERNVTRY 295
 Qy 301 GIAVLGHYLRORDDSSFLREIRITASDPDERFPFNVDAAALNDVIALGRIPLGLBS 360
 Db 296 AVAYLVGYNRRGINDETFLNEIKYIASDPDKHFFNVTDAAKNDIVALGDRIFSLBGT 355
 Qy 361 HAENESSFGLMSQIGFSTRRLKDGILFGMVAYDMGSSVLMLEGHRLFPPRMALDEP 420
 Db 356 N-KNETSFGLEMSQIGFSSHVVEDEGLGAVGAYDMNNAVLKESAGKVIPLRESYLKEF 414
 Qy 421 PPALONHAAYIGYVSSMLRGRRLFLSGAPRRPHRGKVIAPOLKCDGAVVAOSLQGE 480
 Db 415 PEBLNKHGAYIGYVTSVSSROGRVYVAGAPRFHHTKVLFTMHNNRSLTIHQAMRGQ 474
 Qy 481 QIGSYFSGELCPDLPDRDGTDDVLVLAAPMLGPNKTEGKVYVYLVGQGSLLTQGLQ 540
 Db 475 QIGSYFSGELTSDVDGDDGVTDLVVGAMFNP-EGRRERKVVYEL-RQNRFPVYNGTLK 532
 Qy 541 PEPP-QDARFGFAMGALPDNLQDGFADYAVGAPLEDHGQALYLVHGTQSGVRPAPORI 599
 Db 533 DSHSYQANRFGSSIASVADINDQSDYNDVVGAPLEDHNAAGAIYIFHGRGSLIKTPKRI 592
 Qy 600 AAASMPHALSYFGSSVSDGRDLDDGDLVDVAVAGAAIILSSRPVHLTSLSEVTPAI 659
 Db 593 TASELATLQYFGGSIHQDLNEDGLDLVAGVAGNVAIIMSPPVVOINASSLHPEPSKI 652
 Qy 660 SVVORDCRRGOEAVCLUALCPQVTSRTPGRMWHQFMRFTASLDEWTAAGARAFDSG 719
 Db 653 NIFHDCRSRSDATCLAAFLCTPPIFLAPHFQTTVGIRINATWDERRYPRRALDGG 712
 Qy 720 QLSPPRLRLSVGNVTCQQLHFVLDTSYDYLPAALVTFLADNTTKGPVILNBSGSPSI 779
 Db 713 DRFNRAVLSSGGELGERINFHVLDTADYKVPVTFSEVYSLBBDPH-GPMLDGMWPRTL 771
 Qy 780 QKLVPSKDCGPPDNECVTDLVLQVNMDI-----RGRK-----APFV 817
 Db 772 RVSVFPWNGCNEDHCVPDLVLDARSDLPTAMEXCORVLRKPADCSAYTSLPDTVFII 831

Qy	600	AAASPHALSYGRKVDGRDLDDGDDLDVAVAGAGAAIILSSRPVYHLTSPLEVTPOAI	659
Db	593	TASELATGIQVFGCSIHQQLDINEDGLIDLAAAGNAVILMSRPVYINASTLHPEPSKI	652
Qy	660	SVVGDCRRRGGEAVCTLAALCFQYTSRPTGWDHQFVWRFPASIDENWTAGARAAPDSG	719
Db	653	NIHFHDCGRSGRDACTLAALFCLFTPIFLAPHQTTVTGIRYATMTDERRYTPRAHDEG	712
Qy	720	QRLSPRLRLVGNVTCEQLHFHVLDTSDIYLRPAVALTYTFALDNTTKPGVPLEGSPSI	779
Db	713	DRFTRAVILSSGQELCERINPHVDIADYKVPVFSVEYSLHEDPDH-GPMLDDGMPTTL	771
Qy	780	QKLVESFSDCCGPDNCCVTDVYLQVNMMDI-----RGRK-----APRV	817
Db	772	RVSVFPMNGCNEDEHCVDELVDLARSDLPTAMEYQFRLRKPAQDCSAVYLSFDTTVPII	831
Qy	818	RGRGRKAVYSTLERKKEKNAVTSLSIIFSRMLHASTLPQRESEPIKECAAPS--AAAR	875
Db	832	ESTRQRAVEATLEKGENAVSTVINISQSANLQFASLIQKDSQSGSIECVNEERLQKQ	891
Qy	876	LCSVGHVPVQTKAKYTFLLSEFFSCSSLLSQVFGKLTASDSLEKNGTLQENTACTSAVI	935
Db	892	VCNVSYPEFPRAKAKAVAFRLDPEFSKSIFLHHEIELAAGSDSENERDSTKEDVAVLPRHL	951
Qy	936	QYEPHLLSSBESTLRYEYHPYGTLP--VGPEPEKTKTLRVQNLCCYVSGILIALPAA	993
Db	952	KYEAQVLEFTRSSLSHVEKVPMSLERYDGIQPPSCIFRIQNLDFPHGMKMTIPI	1011
Qy	994	VAHGNGVFLSSQVLTN--NASC-LVQNLTEPGRPVHPEELQHTNRNGSNTQCOVVR	1050
Db	1012	ATRSNGRLKLKRLDFTIDEANTSCNMGSTERYPRPVE-EDLRARPQLNHSNDVVSINC	1070
Qy	1051	HLGOLAKGTEVSGLLRLVHNEFFPRAKFKSLTVVSTELGTEGSGVLQLTASRWSESL	1110
Db	1071	NI-RLVPMQEIHFILGLMLRSLKALKYKSMKIMVNAALQRFSPFIREFDEPSQIV	1129
Qy	1111	LEVQOTRILISLWLLIGSVGGILLALLVPCMLKGLPEFAKKIPEEKREKLE	1166
Db	1130	FEISKQEDWQVPIWITGVSTLGGILLALLVLAIAWKLPFNSAR---RRRPPGLD	1181
RESULT 8			
AAU10551			
ID	AAU10551 standard; Protein; 1188 AA.		
AC	AAU10551;		
XX			
DT	14-FEB-2002 (first entry)		
XX			
XX	Human A259 polypeptide.		
XX			
KW	Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;		
KW	liver disease; fibrosis; lung; kidney; bone associated disorder; blood;		
KW	cartilage associated disorder; haematopoietic disorder; bone marrow;		
KW	immune related disease; apoptotic disorder; neuronal tissue disease;		
KW	neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;		
KW	nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;		
KW	antiarthritic; antianaemic; antiallergic; antisthmatic; dermatological;		
KW	antidiabetic; anticonvulsant; antiparkinsonian.		
XX			
XX	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	Peptide	1..22	
FT		/note= "Signal peptide"	
FT	Domain	1..1141	
FT		/note= "Extracellular domain"	
FT	Protein	23..1188	
FT		/note= "Mature human A259"	
FT		37..90	
FT	Domain	/note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"	
FT			

FT	.Domain	115..157 /note= "Integrin alpha repeat domain"
FT	.	164..345 /note= "I domain or Von Willebrand Factor type A domain"
FT	.Domain	367..392 /note= "Integrin alpha repeat domain"
FT	.Domain	421..472 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
FT	.Domain	476..532 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
FT	.Domain	538..593 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
FT	.Domain	600..654 /note= "Integrin alpha repeat domain. The specification states that this domain exists in human A549"
FT	.Domain	1142..1164 /note= "Transmembrane domain"
FT	.Domain	1165..1188 /note= "Cytoplasmic domain"
PX	WC2001B1414-A2.	
PD	01-NOV-2001.	
PF	27-APR-2001; 2001WO-USJ3516.	
PR	27-APR-2000; 2000US-O561263.	
XX	(MILL-) MILLENNIUM PHARM INC.	
PI	Pan Y, Lora J;	
DR	WPI; 2002-041397/05. N-PSDB; AAS16873.	
PT	New A259 nucleic acids and polypeptides, which comprise integrin alpha subunit, useful for diagnosing, preventing or treating e.g. liver disease, kidney or lung fibrosis, cancers, blood disorders or immune related diseases -	
XX	Claim 9; Fig 1; 168bp; English.	
CC	The invention relates to human and murine A259 nucleic acid molecules which encode secreted proteins with homology to Integrin alpha subunits, specifically to integrin alpha 10. The A259 polypeptide and nucleic acid are useful for treating liver disease or fibrosis, particularly kidney fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also useful for diagnosing, preventing or treating cartilage and bone associated disorders (such as bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and osteoporosis), bone marrow, blood and haematopoietic disorders (such as acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune related diseases (such as HIV, viral infections, cancers, T cell autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g. asthma and psoriasis), apoptotic disorders (such as systemic lupus erythematosus and insulin-dependent diabetes mellitus), diseases of the neuronal tissues (such as epilepsy and muscular dystrophy) and neurodegenerative diseases (such as Parkinson's disease and Huntington's disease). This sequence represents the human A259 polypeptide.	
SQ	Sequence 1188 AA;	
OY	Query Match 40.0%; Score 2439.5; DB 23; Length 1188; Best Local Similarity 43.2%; Pred.No. 7.5e-208; Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16	
DG	1 MELPVTHTLPLDVLTLGLCSFNVLDHHPRLPGCPBAEFGYSVLQHVGSGQRMTLVGA 60 :: :: :: :: :: 1 MDLPGRGVVAALSLTMGFDTTFNMDDPKRPKPVIGSRTAFEGTYVOOHDSIGNKWLVYGA 60	

QY	61	PWDGSGRRGGRGVNYCPCVGCAGNAPCAKHLDDYOLGNSHBAVNMHLGMSLLETTDGGG	120
Db	61	PLETNGYOKTGVYKCPV--IHGNTKLNLRVLTLSNVSEKDMRLGLSLATPXDNS	117
QY	121	FMACAPLWSRACGSSVFFSSGICARVDASFOPQSLAPTAQRCPTMYDVIVLDSGNSIYP	180
QY	118	FLACSPPLMSHECGSSYYTTGMCSSRVNSNFRFSKTVAPALQRCQTYMDIYIVLDSGNSIYP	177
QY	181	MSEVOTFLRLVGLKFLPDEPOI QVGLVOYGESPYHMSIGDPRTEEVYRAKNSRREG	240
Db	178	WVEVQFLINLIKKEYIGGOIQOVGVQGEDEVHNEFLINDRVSVDVDEASHIEQRGG	237
QY	241	RETKQAOLIMVACTGSEFSGSHGSRPEALALVVNTDGESSHDEBEPALAKACEAGRVTY	300
Db	238	TETRTAFGLIEFARSEAFQK--GGRKAKKVMYIITDGESHDPDEKVIQOGERDVTYR	295
QY	301	GIAYLGHYLRQRDRDSSFLREIRITIASDEDERFFENVTDDEALTIDVALDGRIFGLEGS	360
Db	296	AAVALGYNNRGINNETFLNEIKYIASDPDDKHFNVTDEALAKDIDVALDGRIFSLBET	355
QY	361	HAENESSGLEMSQIGFSTHRLKDGILFGMVQAYDWGGSVLMLEGGHRLFPRMALDEEF	420
Db	356	N-KNETSGLEMSQIGFSSHVVEDGVLGAVAYPMNGAVLKETSAKAVIPLRESYLKEF	414
QY	421	PRALONHAUYLGYSVSMRLRGRRFLPSGARGPRHGRGVIAFOUKQOAVVAOSLOJE	480
Db	415	PEELNKHGAYLYIYTVSVSSRQGRYYAAGAFENHTGVLLFTNNHNSLTITHQMRQ	474
QY	481	QIGSYFSGELCPLDPRDQDTTDBVLLVAAPMFAPQNKETGRVYVVLVGOOSLTLTGTLQ	540
Db	475	QIGSYFSGEITSDVIDDQGVTDVLLVGAQMYEN--EGREBKGYVVEL--RQNFVYVNGTLX	532
QY	541	PEBP--QDARFGRAMGALPDJNDGPRADAVAGAPLEDGHOGALYLYHGTOSGVRRPHQRI	599
Db	533	DSHSYQNAARFGSSIASVRLNDQSDYNDVVGAAPLEBNHAGALYIFHGFGSSILKTPKORI	592
QY	600	AAASPHMLSYGRGVNDGRLDIDGGDVLVVAAGAGAAIILSSRPVYHLPLELVTPQAI	659
Db	593	TASELATQIYFGCSIHGQLDINEDQLIDLAVGALGNAYILMSRPVQJNASHLEBPSKI	652
QY	660	SVVQDCRRRQGEAVCLTALCFQVTSRTPRGAMHQFYVRFASJLDEWTAGARAAFPDGG	719
Db	653	NI FHRDCKRSGRDAICLAFICFTPI FLAPHQTTVGIRNAYATMDERITYTPRAHIDEG	712
QY	720	QRLSPRLRLSVGNVTCBQLAHFVLDTSYLRPVALTVPALDNTTKPGPVINEGSPISI	779
Db	713	DRFTRAVILSSGOELCERINFHVLDIADYVVPFVSYSLEDBDH--GPMDDCWPITL	771
QY	780	QKLVFSSDCGPDNCCVTDIVYQVMMDI-----RGRK-----APFV	817
Db	772	RVSVFPMGNCNDEHCVPDVLVDABSDLTAMEYCORVLRKRAQDCAVYTSFDTTVPFI	831
QY	818	RGRGRKVLVSTTLERNKEVAYNTLSIIPSRMLHLASLTPQRESPIKVECAPS--AHAR	875
Db	832	ESTRVRVAEALLENKGENAYSTVUNISSALQFASLQKDSOOSIECVABEERLQKQ	891
QY	876	LCSVGHVPVQTAGAKVTLLEBEFFSCSSLSIQVFGKLTASDSJLERNGLQENTAOQASVI	935
Db	892	VCNVSYPEFRAKAKVAFRLDEFFSKSIFLHLEIEILAAQSDNERDSTKEDVAPRFL	951
QY	936	QYEPHLLSSSEBTLKRYEYHRYCTLR--VGRPEFTTLRYONLGCYVVSGLIISALPA	993
Db	952	KYEAADVLEFTRSSLSLHYEKKPSSLERYDGIAPPSCIFRIQNLGIFPHGMMKITIPI	1011
QY	994	VAHGGNYFLSIOVNTN--NASC--IVONTEBPGRVHBELOHNRNLGNSVTCQOVARC	105
Db	1012	ATRSGRNLLKLKRLDELDEANTSCNIGWSTERTPRPVE--EDLRARQLNHSNDVYSTNC	107
QY	1051	HLGOLAKGTEVSGILRLVHNEFFRRAKFXSLTVVSTEBLGTBEGSVIOLTEASWSESL	1110
Db	1071	NI-RLVPRQEIINFHLGLNLMRLSLKALKYKSKIMVNALNQPHSPFIIFREDDPSRQIV	1122
QY	1111	LEVOTRPLLSMLLIGSVLGGLLLLALLVCLMKLGEFAHAKKIPEEKREBKE	1166

```

Db      1130 FEISKQEDNQVPIWIIIVSTIGULLLALVLVALLMKIGFFPSAR----RRREGLD 1181
RESULT 9
AAU14467
ID  AAU14467 standard; Protein; 1188 AA.
XX
XX  AAU14467;
DE
DE  24-OCT-2001 (first entry)
XX
XX  Human novel protein #338.
XX
XX  Human; novel protein; Antinaemic; osteopathic; antiinflammatory;
KM  immunomodulatory; cytosstatic; neuroprotective; vulnery; nocitropic;
KM  anticonvulsant; antiaerthetic; cerebroprotective; antifungal; antiviral;
KM  antihypertic; antiallergic; dermatological; haemostatic; antiasthmatic;
KM  thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KM  Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX  tissue regeneration; immune disorder.
XX
XX  Homo sapiens.
XX  OS
XX  WO200155437-A2.
XX  PN
XX  02-AUG-2001.
XX  PD
XX  25-JAN-2001; 2001WO-US02623.
XX  PF
XX  25-JAN-2000; 2000US-0491404.
XX  PR
XX  (HYSE-) HYSEQ INC.
XX  PA
XX  Tang YT, Liu C, Drmanac RT;
XX  PI
XX  WPI: 2001-451939/48.
XX  DR
XX  N-PSDB; AAS22772.
XX  DR
XX
XX  Isolated polypeptides useful for treating anti-inflammatory diseases,
PT  nervous system disorders, and for regenerating bone and cartilage -
PT
XX
XX  Example 4; Page 828-831; 894pp; English.
XX  PS
XX
XX  The invention relates to polynucleotides encoding novel human
CC  proteins or their active domains. The polypeptides, polynucleotides and
CC  antibodies raised against the polypeptides are used in a method of
CC  treatment of a mammal and prevention of disorders caused by the aberrant
CC  protein expression or activity. The polypeptides can be used as
CC  molecular weight markers, food supplements, and in antibody production.
CC  The polypeptides are used to identify compounds which bind to the
CC  polypeptides. Polynucleotides of the invention are used as probes and
CC  primers, for sequencing, for chromosome or gene mapping, in the
CC  production of recombinant proteins, and in generating anti-sense DNA or
CC  RNA and in gene therapy. Polypeptides of the invention can be used to
CC  target drugs to a tumour, in assays to determine biological activity, to
CC  raise antibodies/elicite an immune response, to determine quantitative
CC  protein levels, as tissue markers, and to isolate receptors or ligands.
CC  Polypeptides of the invention may also be useful in treating platelet
CC  disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC  ligament and/or nerve tissue, wound healing, treating burns, promoting
CC  the proliferation, differentiation and survival of stem cells, as a
CC  contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC  Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC  sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC  fungal infection or from autoimmunity, cancer, allergy, asthma,
CC  graft-versus-host disease, eczema, hemophilia, thrombosis.
CC  anti-inflammatory diseases, nervous system disorders, and infection.
XX  The present sequence represents a protein of the invention.
XX
XX  Sequence 1188 AA;
XX

```


Best Local Similarity 43.2%; Pred. No. 9.2e-208;
Matches 517; Conservative 213; Mismatches 421; Indels 45; Gaps 16;

QY 1 MELPFTNHLFLPLVFLTGCPSPFNDEHHPLPFGPPEAFPGSVYQHVGGRWMLVGA 60
1 MDLPKGLVAVMALSLMPGFTDTFNMDTKPRVIPSRTAFPGYVQOQHISGNKMLVGA 60
Db 1 MDLPKGLVAVMALSLMPGFTDTFNMDTKPRVIPSRTAFPGYVQOQHISGNKMLVGA 60
QY 61 PWDGSGDRGDVYRCPCVGANAPCAKGLGDYQUGNSHPAVMHLGMSLLETDDGCG 120
61 PLETNGYQKTGVYKCPV---IHGNTKLNLRVTLNSVSEKDNMRGLSLATPKDMS 117
Db 61 PLETNGYQKTGVYKCPV---IHGNTKLNLRVTLNSVSEKDNMRGLSLATPKDMS 117
QY 121 PMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQCPTVMDVIVLDGNSIYP 180
121 PMACAPLWSRACGSSVFSSGICARVDASFQPGSLAPTAQCPTVMDVIVLDGNSIYP 180
Db 118 FIACSPFMSHECGSSYYTGMCSRVNSNRFKTVAPALQRCQYMDIYIVLDGNSIYP 177
118 FIACSPFMSHECGSSYYTGMCSRVNSNRFKTVAPALQRCQYMDIYIVLDGNSIYP 177
QY 181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOYGESVHEWMSLGDRTKEVVRAKNSRRRG 240
181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOYGESVHEWMSLGDRTKEVVRAKNSRRRG 240
Db 178 WVEVQHFLINILKKFYIGPGQIQGVGVYGEDVHEFHLDYRSVDVVEASHLEQRCG 237
178 WVEVQHFLINILKKFYIGPGQIQGVGVYGEDVHEFHLDYRSVDVVEASHLEQRCG 237
QY 241 RETKTAOAIMVACTEGFSQSGHGRPEARLLVVTGDSHDEBELPALKACEAGRVTRY 300
241 RETKTAOAIMVACTEGFSQSGHGRPEARLLVVTGDSHDEBELPALKACEAGRVTRY 300
Db 238 TETRTAFGLEFARSEAFQK--GGRKGAKKVMIVITGSHSDPDLKVIQSEERDVTRY 295
238 TETRTAFGLEFARSEAFQK--GGRKGAKKVMIVITGSHSDPDLKVIQSEERDVTRY 295
QY 301 GIAVUCHYLRQRDPSSFLREIRTTASDDEPFENVTDALTDIVLAGRITGLGSG 360
301 GIAVUCHYLRQRDPSSFLREIRTTASDDEPFENVTDALTDIVLAGRITGLGSG 360
Db 236 AAVAVIGYVRRGINPETFLNEIKYIASDDDDKGFENVTDALAKDIVDAGRITSLGEGT 355
236 AAVAVIGYVRRGINPETFLNEIKYIASDDDDKGFENVTDALAKDIVDAGRITSLGEGT 355
QY 361 HAENSEFGLMSQIGFSTRHLKDGILFGMVAGYMDGSLVMBEGHRLFPFRMLADEGF 420
361 HAENSEFGLMSQIGFSTRHLKDGILFGMVAGYMDGSLVMBEGHRLFPFRMLADEGF 420
Db 366 N-KNELTSFELEMSQIGFSSHVVEDGVLLGAVGYMDNGVNLKETSAGKIPILRESTLKEF 414
366 N-KNELTSFELEMSQIGFSSHVVEDGVLLGAVGYMDNGVNLKETSAGKIPILRESTLKEF 414
QY 421 PPALONHAAYLGVSYSSMLRGRRLLFSGAPFRHGRGVIAFOLKDGAVVAOSLQGE 480
421 PPALONHAAYLGVSYSSMLRGRRLLFSGAPFRHGRGVIAFOLKDGAVVAOSLQGE 480
Db 415 PEELNKHGAYLGLYTVTSVSSRQGRVYVAGAPRPHNTGVILFTMHNNNSLTIHQAMRQ 474
415 PEELNKHGAYLGLYTVTSVSSRQGRVYVAGAPRPHNTGVILFTMHNNNSLTIHQAMRQ 474
QY 481 QIGSTFGEELCPLDTRDGTDTVLLVADPMFLGPQNKETGRVYVYLVGQSLLTLQGTIQ 540
481 QIGSTFGEELCPLDTRDGTDTVLLVADPMFLGPQNKETGRVYVYLVGQSLLTLQGTIQ 540
Db 475 QIGSYFGESEITVDIDGDTVDVLLVGAEMYPN--EGREGRKYYVEL--RQNRFFVNGTLK 532
475 QIGSYFGESEITVDIDGDTVDVLLVGAEMYPN--EGREGRKYYVEL--RQNRFFVNGTLK 532
QY 541 PPPP--ODARFGFAMGALPDLNODGFADVAVGAPLEGHGALYLVHGTSQGRPHPAQCI 599
541 PPPP--ODARFGFAMGALPDLNODGFADVAVGAPLEGHGALYLVHGTSQGRPHPAQCI 599
Db 533 DSHSYONARFGSSIASVRDLNODSYDVVAGAPLEBNHAGAIYIFHGFSGSLTKPKQCI 592
533 DSHSYONARFGSSIASVRDLNODSYDVVAGAPLEBNHAGAIYIFHGFSGSLTKPKQCI 592
QY 600 AAASMPHALSYFGRSVGRLDLDGDLVAVAGAOGAAILSSRPVHLTPSLLEVTPQAI 659
600 AAASMPHALSYFGRSVGRLDLDGDLVAVAGAOGAAILSSRPVHLTPSLLEVTPQAI 659
Db 593 TASELATGQYFGCGSIHQGLDINBGLIDLAVALGALNAILMSRPVQVQINASHPEPSKI 652
593 TASELATGQYFGCGSIHQGLDINBGLIDLAVALGALNAILMSRPVQVQINASHPEPSKI 652
QY 660 SVVQDRCRRRGEAVCLTALCFQVTSRTGPRWDHGFYMRFTASLDEWTAGARAAFDGSG 719
660 SVVQDRCRRRGEAVCLTALCFQVTSRTGPRWDHGFYMRFTASLDEWTAGARAAFDGSG 719
Db 653 NIFHDDCKKSGRADATCLAAFLCFTTPIFLAPHEQTTVGIRYNAVIMDERYTPRAHDEGG 712
653 NIFHDDCKKSGRADATCLAAFLCFTTPIFLAPHEQTTVGIRYNAVIMDERYTPRAHDEGG 712
QY 720 QNLSRRRLRLSYGNVTCBQHFHVLDTSYLKPVALTIVFALDNTTKPKPVUNESPSI 779
720 QNLSRRRLRLSYGNVTCBQHFHVLDTSYLKPVALTIVFALDNTTKPKPVUNESPSI 779
Db 713 DFTNRKAVLLSSGOELCERINFHVLDTADYKRVFSPVSEVSLDDPDH--GPMLLDDGPTTL 771
713 DFTNRKAVLLSSGOELCERINFHVLDTADYKRVFSPVSEVSLDDPDH--GPMLLDDGPTTL 771
QY 780 QKLVFESKXDCGPDNCEVTDLVQVMMDI-----RGSRK-----APFVY 817
780 QKLVFESKXDCGPDNCEVTDLVQVMMDI-----RGSRK-----APFVY 817
Db 772 RSVSEVFMWNCNDEHCVPLVDARSDLPTAMEYQCRVLRKRAQDCSATLISFDTVFI 831
772 RSVSEVFMWNCNDEHCVPLVDARSDLPTAMEYQCRVLRKRAQDCSATLISFDTVFI 831
QY 818 RGGRRKVLVSTLLENKENAVNTSLISFRNLHLASTLPQRESPIKYCAAPS--AAAR 875
818 RGGRRKVLVSTLLENKENAVNTSLISFRNLHLASTLPQRESPIKYCAAPS--AAAR 875
Db 832 ESTROKVAVEATLENKGENAVSTVINISQSANLQFASLLQKSDSDSGSIECVNEERLQKQ 891
832 ESTROKVAVEATLENKGENAVSTVINISQSANLQFASLLQKSDSDSGSIECVNEERLQKQ 891
QY 876 LCSVGHVPOFGAKTFLLEFEFSCSLSQYFGKLTASDSELENGTLQENTAQTSAXI 935
876 LCSVGHVPOFGAKTFLLEFEFSCSLSQYFGKLTASDSELENGTLQENTAQTSAXI 935
Db 892 VCNVSIPEPRAKAKAFRLDPEFSKSIPLHLEITELAAOSDSENERSTEDVAVPRFL 951
892 VCNVSIPEPRAKAKAFRLDPEFSKSIPLHLEITELAAOSDSENERSTEDVAVPRFL 951
QY 936 QYEPHLFSESESTLHREYVPGTLP--VGPGEEFTTLRVONTGCVVSGHILSALLPA 993
936 QYEPHLFSESESTLHREYVPGTLP--VGPGEEFTTLRVONTGCVVSGHILSALLPA 993
Db 952 KKEVDLFLTRSSLSHVEKPKNSLERVYQIGPFECIRIQLGFIHGMKMTIFI 1011
952 KKEVDLFLTRSSLSHVEKPKNSLERVYQIGPFECIRIQLGFIHGMKMTIFI 1011
QY 994 VAHGGNYFLSLSQVITN--NASC-IYQNTTEPPGPVHBEELQHTNRLNGSTQCQVVC 1050
994 VAHGGNYFLSLSQVITN--NASC-IYQNTTEPPGPVHBEELQHTNRLNGSTQCQVVC 1050

Db 1012 ATRSGNRLKLKRLDTDEANTSCINWGSTERYPTPE--EDLRAPQLNHSNDVVSINC 1070
QY 1051 HLGLQAKTEVSGVGLRLVHNEFPFRAPKSLTVSTPELGTBEGSVQLTEASRWSBL 1110
1051 HLGLQAKTEVSGVGLRLVHNEFPFRAPKSLTVSTPELGTBEGSVQLTEASRWSBL 1110
Db 1071 NI-RVPRQELNFIHLGLMLRSALKKTKSMKIMVNAALORQSPITFREDSDRQIV 1129
QY 1111 LEVOTPRILSLMILIGSVLGLLLALLVFCMLKIGFPAHKKIPEEKREKLE 1166
1111 LEVOTPRILSLMILIGSVLGLLLALLVFCMLKIGFPAHKKIPEEKREKLE 1166
Db 1130 FEISQEDMQVPIMIVISTGLLLALLVIALMKLGFPSAR---RRREPGD 1181
1130 FEISQEDMQVPIMIVISTGLLLALLVIALMKLGFPSAR---RRREPGD 1181

RESULT 10
AAB25582
ID AAB25582 standard; Protein; 1189 AA.
XX
AC AAB25582;
XX
DT 21-NOV-2000 (first entry)
XX
DE ITGA11 protein encoded by human secreted protein gene #7.
XX
KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
KW anticancer; vulnery; antiviral; antibacterial; antifungal;
KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
KW Crohn's disease; nephritis; hyperproliferative disorder;
KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
OS Homo sapiens.
XX
PN WO200029435-A1.
XX
PD 25-MAY-2000.
XX
PF 27-OCT-1999; 99WO-US25031.
XX
PR 28-OCT-1998; 98US-0105971.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;
PI Greene JM;
XX
DR MPI: 2000-387742/33.
XX
DR N-PSDB; AAA80612.
XX
PT Isolated nucleic acid molecules encoding human secreted proteins are
PT used for the prevention, amelioration and treatment of autoimmune,
PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
PT wounds, and infectious diseases -
XX
PS Claim 1; Figure 19A-F; 803pp; English.
XX
XX
XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given
XX in AAB80606-A80623 encode the 12 secreted protein sequences given in
XX AAB25576-B25593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant;
XX anti-inflammatory; antirheumatic; antirheumatic; dermatological;
XX antiproliferative; antiarteriosclerotic; anticancer; vulnery;
XX antiviral; antibacterial; and antifungal activity. The proteins,
XX polypeptides, agonists and antagonists may be used to treat prevent
XX and/or diagnose various disease, disorders and conditions examples of
XX which include: immune disorders e.g. Addison's disease, rheumatoid
XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
XX hyperproliferative disorders such as paraproteinemias and purpura;
XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
XX sequences may also be used in wound healing and the treatment of

CC Infectious diseases. The human secreted protein gene #7 and protein
 CC sequences are represented in sequences AAA80612 and AAA25582. Secreted
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
 CC AAA80652-AB80661 represent genes which are related to the secreted protein
 CC gene#7.
 CC
 XX
 SQ Sequence 1189 AA;
 Query Match 39.8%; Score 2429; DB 21; Length 1189;
 Best Local Similarity 43.1%; Pred. No. 6.5e-207; Indels 46; Gaps 16;
 Matches 516; Conservative 213; Mismatches 422;
 QY 1 MELPFTLHFLPLVLTGLCSPEFLDEHRLPPGPEAEFGYVLQHWGGGRMLVGA 60
 Db 1 MDLPRLVAVAMALSLWPGFTDFTNMOTRKPRVIGSRTAFYGTVOQHDSIGKMLVGA 60
 QY 61 PMDPSGDRGDYRCPVGGANAPCAKHLGDIYQIENSHPAVNNHLSLETGGD 120
 Db 61 PLETNGYOKTGDYVKKCPV--IHGCTKMLGRVTLISNVSERKDNRLGLSLATNPKDMS 117
 QY 121 FMACAPLWMSRACGSSVFSGICARVDASFOQSLAPTORCPTVMDVILDGNSIYP 180
 Db 118 FLACSLWMSHECCSSYTTTGMCSRANSNFRSKTVAPALORCOTYMDIYVLDSNSIYP 177
 QY 181 WSEVQTLRLVGLKFLIDPEQLQVGVQGESPVHMSLDFTKEEVYRAAKNLAREG 240
 Db 178 WVEVQHFLINILKKFYIGPQIQGVQGEVDVHEFLINDVSVADVVEAASHIEQGG 237
 QY 241 RETKTQALINACTEGFSQSHGGRPEARLLVVTGTGSEIDGSELPAALKACAGVTRY 300
 Db 238 TETRTAFGIEFARSEAFQK--GGRKAKKVMYITGESHSDSDLEKRVQOSRDVTRY 295
 QY 301 GIAVLGHYLRQRPDPSFLREIRTIASDPDERFFPVNTEBAALTDIVDALGRIEGL 360
 Db 296 AVAVLGYNRGNGPTEFLNEIKYIASDPDDKAFVNTDEBAALKDVIDALGRIEGL 355
 QY 361 HAENESSFGLMSQISFSTRKLGKILFGVAVGADGCVLMEGRLLPFRMALEDEF 420
 Db 356 N-KNETSFGLMSQTFSSHVEDVLLGAVGADWNGAVLKETSAGKIPLAESYLEKEF 414
 QY 421 PPLQNHAAVLGYSVSSMLRGRRLPLSGAPFRFRKGVIAQLKKDQAVVAQSLQSE 480
 Db 415 PEBLKNHGAVLGYTVSVSSRGRVYVAGAPRFNHTGKILFTMHNNSSLTIHQARQ 474
 QY 481 QIGSYGSELCPTDTRDGTDTDLVLAAPWFLGPOKKEGRVYVYLVGQSLTTLTGTO 540
 Db 475 QIGSYGSEITSDIDGDTVDLVLAGAPWYFN-EGREKGVYVEL-RQNRVYVGTIK 532
 QY 541 PEBP-ODARFGFAMGALPDLNODGADVAVGAPLEDDHOGALYLYHGTOSGVAPHPAORI 599
 Db 533 DSHSYQNAFFGSSIASVRDNLGDSYNDVVYGAFLLEDNHAGAIYIFHGFSGSLIKTEKORI 592
 QY 600 AAASMHALSYRPSVDGRLLDGDLDLVVAAGCAAILLSRPIYHLTPSLSEVTPQAI 659
 Db 593 TASELATGQYFCGSIHQGLDNLDEGLIDLAVAGLNAVILMSRPVQVQINASHPEPSKI 652
 QY 660 SVVQRCRRGCGEAVCTALCFQVTSRTPGCRWDHGFYMRFTSLDEMTAGARAADGSG 719
 Db 653 NIFHRCKSGRATLAAFLCTPTPLFAPHFQTTVGIRYKNTMERKRYTPRAHLDEGG 712
 QY 720 QRLSPRLRLSVGNVTCEQLHFVLDTSVLRVALVTYFALDNTTKPGVLENGSPISI 779
 Db 713 DRFTNAVLVSSGQELCERINFHVLDTADVKKVFTSVESYLEDPDH-GPMLDDGPTTL 771
 QY 780 QKLVPSKQCGPNECVTLVLOVMDI-----RGRSK-----APFV 817
 Db 772 RVSVPFWNGCNEDEHCVPLVLDARSDLPAMEYCORVLARKPAQDCSAYTLSEFTVFI 831
 QY 818 RGRGRVAVLSTTDEKNAVNTSLISFRNLHLASLTPQRPSPKVECAAPS-AAAR 875
 Db 832 ESTRQAVAVATLNGENNAYSVLIWISGANIQFSLQKEDSDSIECVNEERLQKO 891
 QY 876 LGSVGHVPQTGAKVTFLEFEFSCSSLSQVGVKLTASSDSLERNGLQENTQTSAYI 935

Db 892 VCNVSYPPFRKAKAVAFRIIDFEFSKSIPLHHEILAGSDSNEDSTKEDVABLRHL 951
 QY 936 QYEHLLHSSSESTLRHYEVHPTLP--VGPOPEKTLIRYONIGCVYSGIIALPA 993
 Db 952 KYEADVLFTRSSSLHYEVKLNSSLERYDGIQPPSCIFRIQNLPLPHIGIMKITPI 1011
 QY 994 VAHGNVFLASLSQVLTN--NASG-IYONLTPEPPVAPPELQHTNRNGSNTOCQVR 1049
 Db 1012 ATRSGNRLKLKRLDPLDEVANTSCNIMGNSTYRTPYE-EDLRAPQLNHSNDSVAIN 1070
 QY 1050 CHLGQLANGTEVSVGLRLVHNEPFRKPKSLTVVSTFELCTEEGVSVLQLTASRWSES 1109
 Db 1071 CNI-RLVPMQEIINFHLGNLMLRSIKAKYKSMKIMVAAAROPHSEFIFREEDPSRQI 1129
 QY 1110 LLEVQTRPILSLWILGVSVLGILLALLVFCIMKGFFAHKKIPEEKREKLE 1166
 Db 1130 VFEISKQEDWQVPIWIVGSTIGLLALLVALLMLKLGFFRSAR---RRREPGLD 1182
 RESULT 11
 ABG12949
 ID ABG12949 standard; Protein; 1189 AA.
 XX ABG12949;
 AC
 XX
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #12940.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS77136.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 43308; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_sequences.

XX Sequence 1189 AA;

Query Match 39.8%; Score 2429; DB 22; Length 1189;
 Best Local Similarity 43.1%; Pred. No. 6,5e-207;
 Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

```

QY 1 MELPVTHFLRLVLTGLCSPPNIDHHRLLPGRPEAFGYSVLQHGSGQRMVLYGA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MQLPGLVYAMALSLMPGFTDFNMMDTRKPRVYPSGRFAFGYVQGHIDISGNKMLVYGA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 PNDGSGDRGDRVYRPPVGAHANAPCAKHLDYQLGNSHRAVNMHLSLLEDDGG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 PLETNGYQKTGVYKCPV---IHGNTKLTNLGRVTLSNVSEKDNMRLLSLATNPKNDS 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 FMACAPLMGRACSSVFSVSGICARVDASFQPOGSLAFTQRCPTVMDVIVLDGNSIYP 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 118 FLACSPLMHSECGSSYTTTGMCSRNSNFRPSKTVAPALQRCQTYMDIYIVLDGNSIYP 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 WSEVQTFRLRVGKLFIDPEQIQVGLVYQGESPVHEMSLGDPRTEKEVRAAKNLSRREG 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 178 WVEVQHFLLNLIKFKYIGQIQVGVQYGEDVVEHFLNDRSVADVVEASHIEQRG 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 RETKTQALIMVACTEGFSQSHGPRPAARLLVVTVDGSHDGBELPALKACBAGRTRY 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 238 TETRTAFGLFEPARSEFQK--GGRKGAKKMIVITDGBSHDSDPLEKVIQGBERDVTY 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GIAYIGHYVRRODPSPFLREIRITIASDPERPFVNTDEALITDIDALGDRIFLEGS 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 AVAVLGYNRKGINPFTFLNEIKYIASDDDKFFNVTDALKIDVIDALGRIFSLBET 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 HAENESSFGLMSQIGFSTHRLKDGLFGWGAIDWGGSVLWEGHRLFPPRMALDEDF 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 N-KNETSFGLMSQIGFSSHVVEDVLLGAVGYDMNGAVLKEYSAGKVIPLRESYLKEF 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 PALQNHAYLVGSYSMLRGRLFLSGARFRFRKGVIAFQKKQDAVVAQSLQSE 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 PEELKNHAGLYTVIVSVSSRQGRVYVAGAPRFNHTGVILFTMMNNNSLTQHQAQ 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 QIGSYFGEELCPIDRDGTTDVLLVAAMPFLGPQNKETGRVYVYLVGQSSLLTQGTQ 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 QIGSYFGEELTSDVDGDTVDVLLVGAEMVFN--EGREKGVYIEL--RQNRVYINGILK 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 PEPP--QDAFFGAMGALPDINODGFADVAVGAPLEDHOGALYLYHGTQSGVRPHPAQR 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 533 DSHSYQNAFGSSIASVRLNDQSYNDVVVGAFLBNHAGAIYIPHGFSGSLKTPKQRI 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 AAASMPHALSYGRSVYDRLDLDGDDLVVAVGAGCAALILSSRPVHLTPELEVTPQAI 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 593 TASELATGQYFGCSIHQDLNEDGLIDLAAGALCNAAVILMSRPVQINASLHFEPSKI 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 660 SVVQDDCRRGQEAVALTLALCEQVTSRTPGKMDHFVWRFTASLDEWTAGARAFDGG 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 653 NTFHDDCKRSGDATALFLCFTPIFLAPHOTTTVGIRYNATWDERRYTPRAHIDEGG 712
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 720 QRLSPRLRLSVGNVTCQQLHFVLDTSYLRPVALTVFALDNTTKPGPVLNBSPTSI 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 713 DFTNRALVLLSSGQELCERINHFVLDYAVKVFESVYSLEDPDH--GPMLDGMPFTLL 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 780 QKLVPSKQCGPDNCEYDLVYQVMNDT-----RGRK-----APRV 817
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 772 RVSVPWMNCNDEDEHCVPLVLDARSDLPTAMEYCORVLRKPAODCSATYLSFDTVFII 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 818 RGRARRVLLSTILENKENAVNTSLIIFSRNLHLASLPPRESPIKVCAPPS--AHAR 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 832 ESTRQAVAVATLENGENAYISTVNISSANIQFASLIQKEDSDGSIETCVNERRLQK 891
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 876 LCSVGHVYFQIGAKTFLLEPFSCSSLLSOVFGKLTSSSLSBERNTLQENTQOTSAYI 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 882 VCNVSYPPEFRAYAKAFRLDEPFSSKSIPLHLEIEIAGSDSNSNEBSTDENVAFLRPHL 951
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 936 QYEPHLFSSSSTLHRYVHVHYGLP--VGPGEKTKTLRQVNGCYVVGSLIISALLPA 993
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 952 KTEADVLTFRSSSSLSHYEVKKNLSLERYDGPPECTFRIQNLGLFPHGMKMTITFI 1011
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 994 VAHGNVFLSLSQVITN--NASC-IVQNLTEPPGPVPHPELOHTNRLANGSNTQCQVVR 1049
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1012 ATRSGNRLRLKLRDLTDEVANTSCNMGNSNTEYRTPVE--EDLRAAPQLNNSNDSVAIN 1070
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1050 CHLGQLAGTEVSGCLRVLVHNEPFRRAKFKSLTVSTFELGTEBGSVLQLTASRWSBS 1109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 CNI--RLVNOEINFLLGNLWLSLKALKYKSMKIMVAAALQROPSPFIREEDEPSRQI 1129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1110 LLEVQTPRLISLWILGSLVGLLLALLVFCIMKGFPAHKKIPEEKREKLE 1166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1130 VFEISKQEDQVPIWITVIGSTLGLLLALLVLAWMKGFERSAR---RRREPGLD 1182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Result 12

AAB50087
 ID AAB50087 standard; Protein, 1188 AA.

XX AAB50087;

XX AC AAB50087;

XX DT 19-MAR-2001 (first entry)

XX DE Murine A259.

XX KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;

XX KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;

XX KW rheumatoid arthritis.

XX OS Mus sp.

XX Key

XX Key Location/Qualifiers

XX FT 1..22 /label= Signal_peptide

XX FT 23..1188 /label= Mature_protein

XX FT 1..1141 /label= Extracellular_domain

XX FT 39..74 /label= Integrin_alpha_subunit_repeat_domain_#1

XX FT 115..157 /label= Integrin_alpha_subunit_repeat_domain_#2

XX FT 164..345 /label= Integrin_alpha_subunit_repeat_domain_#3

XX FT 367..392 /label= Integrin_alpha_subunit_repeat_domain_#4

XX FT 421..455 /label= Integrin_alpha_subunit_repeat_domain_#5

XX FT 478..516 /label= Integrin_alpha_subunit_repeat_domain_#6

XX FT 540..575 /label= Integrin_alpha_subunit_repeat_domain_#7

XX FT 602..640 /label= Integrin_alpha_subunit_repeat_domain_#8

XX FT 1142..1164 /label= Integrin_alpha_subunit_repeat_domain_#9

XX FT 1165..1188 /label= Transmembrane_domain

XX FT 1165..1188 /label= Cytoplasmic_domain

XX WO200073339-A1.

XX 07-DEC-2000.

XX 15-MAY-2000; 2000WO-US13262.

XX 28-MAY-1999; 99US-0322790.

XX 27-APR-2000; 2000US-0561263.

XX (MILL-) MILLENNIUM PHARM INC.
 PA Pan Y, Lora JM;
 XX WPI, 2001-041142/05.
 DR N-PSDB; AAC91904, AAC91905.
 XX
 PT Nucleic acid encoding alpha-integrin subunits, useful for treatment and
 diagnosis of fibrosis, e.g. of the liver
 XX
 XX Claim 8, Fig 5, 164pp. English.
 XX
 CC The present sequence is murine integrin alpha subunit, A259 is
 CC homologous with the alpha1 and alpha10 integrin subunits and is
 CC overexpressed in fibrosis. A259 is implicated in regulation of
 CC proliferation, differentiation and/or function of many different cell
 CC types. Inhibitors of A259 activity are useful for the treatment of liver
 CC disease, particularly fibrosis, and also fibrosis in other organs
 CC (specifically lung and kidney). In addition, A259 can be used for
 CC treatment and prevention of cancer, osteoporosis, acute myeloid
 CC leukemia, HIV infection, and rheumatoid arthritis.

XX Sequence 1188 AA;

Query Match 39.7%; Score 2422; DB 22; Length 1188;
 Best Local Similarity 43.4%; Pred. No. 2,7e-206;
 Matches 513; Conservative 207; Mismatches 412; Indels 50; Gaps 15;

18 GLCSFNLDEHHRLEFPPEAEFGYSVLQHVGGGRMMLVGAPWDGSGDRRDYRCP 77
 18 GFDFDTMDTRNRRVAGPSAAFFGYTVOQHDSIGKKMLVVGAPMETNGHOKTGDIYKCP 77
 78 VGGANAPCAKGGHGLGYOIGNSHPAVNMHLGMSLLETDDGGGFMCAPLMSACGSSVF 137
 78 VTQGN--CTKMLNGVLTLSNVSERKDNRLGLSLTNPKNLSFLCSPLMSIECCSSYF 134
 138 SSGICARVDASFPQSGSLAFTAQRCTPTMDVIVLIDGNSIYPSWEVOTFLRLVGLFI 197
 135 TTMCMGRVANSNFPFSKTVAPALORCQTYMDIVLIDGNSIYPSWEVQFLNLLKKFYI 194
 198 DEBOIQVLVOYGESFVHEWSLGDFTKEEVRPAKNSRRREGTFKTAQAIVACTEGF 257
 195 GEGQIQVGIYQYGEDVAHFHLNDYRSVQDVEAASHIEQGGTETRTAFGIEFARSEAF 254
 258 SQSHGGRPEAARLLVVVTGESHDSDELPAALACAGVTRTGVIVLGHVLRQDPSS 317
 255 QR--GGRKAKKVMIVITDGESHDSPLLEKVRQSEKDNVTRYAVALGVYNRGINPET 312
 318 FLREIRTIASDPDERFFNVTDDEAALTDIVDALGDRIFGLESGHAENESSFGLMSQIGF 377
 313 FLNEIKYIASDPDDKHFNVTDDEAALKDIDVALGDRIFGLESGH--KNEISFGLMSQIGF 371
 378 STHRLKDGILFGMVGAVWVGSLVMEGGRLEPPPMALDEDFPPALQNHAAVLGYSS 437
 372,SSHVEDGILGLGAVGAYDMNGAVLKETSAGKVIPIHRESYLKEPPELKNHAAVLGYTVNS 431
 438 MLRGSRRLFLSGAPRRHGRKVIAPQLKKDGAIVRAQSLQSGRQISYSGSELCPIDTOR 497
 432 VVSSRGRRYVAVGAPRRHGTGKILFSMHNHNSLTTHQALRGQISYSGSETTSVDVND 491
 498 DGTDTDLVAAPMFLPQNKETGRVVYVLVGGQSLTLTLOGTLQPEBP--ODARFGFMGL 556
 492 DRTVDLVLGAPRYFS--BGRERKVVYVNL--RQNRVVYVGTGLKDSHSYQNAFGSGIAV 549
 557 PDLNODGPAADVAGAPLEDGHOGALYLYHGTOGCVRPHPAQRITAAASMPHALSYGRSYD 616
 550 QDLNODSYNDVVVAGAPLEDGSHRGALYIFHGFQTNILKKPMORTASBELAPGLQHFEGSIH 609
 617 GRDLUDGDLVAVAGAAAILLSRPVHLPLPSLEVPQMSIVQDRRGGQAVL 676
 610 GQDLNLEDGLVLAVALGNAAVLMARPVQVIMASLHPEPSKINIPHKDCKRGRDATCL 669

QY 677 TAAFCQVTSRTPGEMDHFQWRFTASLDEWTAGARAAFDGSGQRSLSPRLLSVGNVTC 736
 DB 670 AAFLECFIPILAPHPQTATVGIKRYATWDERRYMRRAHLDEGGDQFTNRAVLLSSGSHC 729
 QY 737 EQLHFRVVDTSYLRPAVALTTFALDNTTKRGQPVINESSPSTISQLVPEPSKDCGDNCEV 796
 DB 730 QRINHVVLDIADYVPAVFSVEYSLEBDPN--GPMLDNGMPTTLRVSVFPMNGCNDEHCV 788
 QY 797 TDLVQVNMIDRGSK-----APFVVRGGRKVLVSTLENRK 834
 DB 789 FDLVLDASDLPTAMEYQORVLGRPAQDCSSYTLSPDTTVPILBSTRRAVVALELRNG 848
 QY 835 ENAVNTSLSIIFSRNLHSLTPRESPIKYECAPS--AAHRLCSVGHVPFGTGAKE 892
 DB 849 ENAVAVANTISQSENLQFASLIQKDDONSIECVNEERRLHKVGNVSYPPFRAKAKYAF 908
 QY 893 LIEFFSGSLISQVFGKLTASSDSLEKNGTQENTAQTSAYIQVEPHLLSSSESTLRY 952
 DB 909 RLDPEFSKSVPLNHLQIHLAGSDSHEDSTADDTALLRPLKYEADVLFPRSSLSHF 968
 QY 953 EVHPYGTLPV--GPGPEKTLRYONLGCYVVGSLISALLPAVHAGNYSFLSQVITN 1010
 DB 969 EVKANSLSLESDYDGIQPPNCCYFKVONLGFPIHGVMMKITVPDIATRGGRMLMDPFTD 1028
 QY 1011 --NASC-IVONLTPBPBPVHPEELQHTNRUNGSTQOVVYRCHLGOLAKTEVSGLLR 1067
 DB 1029 QGNTSCNIMGNSTERYSTPTE--EDLSHAPQRNHSNDVSIICNL--RLAPSQETSFYLVG 1086
 QY 1068 LVHNEFFPRAKFSLTVVSTFELGTEBGSVLQLTASMSSESLDEVQTRPILISMLLI 1127
 DB 1087 NMLTSLALKYRSLSKIVNALQRFHSPFLFREEDSRQVTFEISKQEDMQVPIWITV 1146
 QY 1128 GSVLGLLLALLVFCMLKLGFAHKK-----IPEEEK 1160
 DB 1147 GSTLGLLLALLVLAHMKLGFFKSAKRRRPGLCPTKEIK 1188

RESULT 13
 AAU10552
 ID AAU10552 standard, Protein, 1188 AA.
 XX AAU10552;
 AC 14-FEB-2002 (first entry)
 DT
 XX
 DE Murine A259 polypeptide.
 XX
 KW Mouse, A259, integrin alpha subunit, integrin alpha 10; secreted protein;
 KW liver disease, fibrosis; lung; kidney; bone associated disorder; blood;
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;
 KW immune related disease; apoptotic disorder; neuronal tissue disease;
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
 KW antidiabetic; antihypertensive; antiallergic; antisthmatic; dermatological;
 KW antidiabetic; anticonvulsant; antiparkinsonian.
 XX
 OS Mus musculus.
 FH
 FH Key
 FH Peptide
 FT 1..22 Location/Qualifiers
 FT /note= "Signal peptide"
 FT 1..1141
 FT /note= "Extracellular domain"
 FT 23..1188
 FT /note= "Mature murine A259"
 FT 39..74
 FT /note= "Integrin alpha repeat domain"
 FT 115..157
 FT /note= "Integrin alpha repeat domain"
 FT 164..345
 FT /note= "I domain or Von Willebrand Factor type A domain"
 FT 367..392
 FT /note= "Integrin alpha repeat domain"

```

FT Domain 421..455
FT /note= "Integrin alpha repeat domain"
FT Domain 478..516
FT /note= "Integrin alpha repeat domain"
FT Domain 540..575
FT /note= "Integrin alpha repeat domain"
FT Domain 602..640
FT /note= "Integrin alpha repeat domain"
FT Domain 1142..1164
FT /note= "Transmembrane domain"
FT Domain 1165..1188
FT /note= "Cytoplasmic domain"
PN WO200181414-A2.
XX 01-NOV-2001.
XX 27-APR-2001; 2001WO-US13516.
XX 27-APR-2000; 2000US-0561263.
XX (MILL-) MILLENNIUM PHARM INC.
XX Pan Y, Lora J;
XX WPI; 2002-041397/05.
DR N-PSDB; AAS16874.
XX
XX New A259 nucleic acids and polypeptides, which comprise integrin alpha
XX subunit, useful for diagnosing, preventing or treating e.g. liver
XX disease, kidney or lung fibrosis, cancers, blood disorders or immune
XX related diseases
XX
XX Claim 9, Fig 5, 168pp; English.
XX
XX The invention relates to human and murine A259 nucleic acid molecules
XX which encode secreted proteins with homology to integrin alpha subunits,
XX specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
XX are useful for treating liver disease or fibrosis, particularly kidney
XX fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
XX useful for diagnosing, preventing or treating cartilage and bone
XX associated disorders (such as bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
XX osteoporosis), bone marrow, blood and haematopoietic disorders (such as
XX acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune
XX related diseases (such as HIV, viral infections, cancers, T cell
XX autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
XX asthma and psoriasis), apoptotic disorders (such as systemic lupus
XX erythematosus and insulin-dependent diabetes mellitus), diseases of the
XX neuronal tissues (such as epilepsy and muscular dystrophy) and
XX neurodegenerative diseases (such as Parkinson's disease and Huntington's
XX disease). This sequence represents the murine A259 polypeptide.
XX
XX Sequence 1188 AA;
XX
XX Query Match 39.7%; Score 2422; DB 23; Length 1188;
XX Best Local Similarity 43.4%; Pred. No. 2.7e-206;
XX Matches 513; Conservative 207; Mismatches 412; Indels 50; Gaps 15;

```

```

Db 195 GPQGIQVGIYGEDVAHEFLNDIRSVKDVVEAASHIEQRGNETRTAFCIEPARSEAF 254
Qy 258 SQSHGRPREARLLVVTVDGESHDEBELPALKCAEAGRVTRYGIAVLGHYLRORDPSS 317
Db 255 OK--GGRKGAKKMVLITIDGESHSDPLEKTIROSEKDNVRYAVAVLYGNKGINET 312
Qy 318 FLREIRTIASDPDERFFPNVTDEALTDIYDALDRIFGLEGSHAENESSFGLMSQIGF 377
Db 313 FLNEIKYIASDPDDKHFFNVITDEAALKDIDLDRIFFSLBGTN-KNETSFGLMSQIGF 371
Qy 378 STHLKQGIILGCMGAYMGGSVYMLBEGHILFFPRMALBDEFPALONHAAYIGYSVSS 437
Db 372 SSHVEDDILLGAAGAYDMNGAVLKETISAGVIFPRESYLKFPBELKNHAAYIGIYVTS 431
Qy 438 MLNRGRRLFLSGAPRFRHGRKVIAFOLKKGAVRVAQSLQGEQISYFGSELCPLDTR 497
Db 432 VVSSRQGRVYVYAGAPRFNHTGKVLIFSNNRSLTIHOALGEOIGSIFGSEITSVDND 491
Qy 498 DGTIDVLLVAAPELGPONKETGRVYVYLVGQSLTLTQGTLODEPP-QDARFGFAMGAL 556
Db 492 DRVTIDVLLVGAPEWFS-EGREGRKVVYVNL-RQNRFYVNGTLKDSHSYQNAFSGCIASV 549
Qy 557 PDLNODGFADYVAGAPLEDHQALYLYHGQSGVRPHPAQRIAAASPHALSYGRSVD 616
Db 550 QDLNODSYNDVAVGAPLEDHSHGAIYIFHGQTNILKKPMKRIASBELAPGLQHFSGSIH 609
Qy 617 GRLLDGDPLDVAVVAGOGAAILLSRPIVLTSPSEVTPAIVSVORDCRRGOEAVCL 676
Db 610 GQDLNENGLVDLAVGALGNVAVLMARVYVINSLSHEPESKINIFHDCRKNRDAICL 669
Qy 677 TAAICFOVTSRTPGRMWHQFYMRFTASIDENWTAGARAAFDGSGQRLSPRLRLSVGNVTC 736
Db 670 AALCFEIPIFLAPHQATVIGIRYNATWDERVMPRAHLDGGOFTRAVALLSGQSHC 729
Qy 737 EQLHFRVLTDSYLRPAVALTYTPALDNTKRGPELVNBSPTSIQKLVFSDCCGPNDCV 796
Db 730 QRINHFVLTDTADYKPAVAFVSLEBPDN-GPMLDNGMPTLTVSVSPFMNGCNEDEHCV 788
Qy 797 TDVLVQVMDRGRSRK-----APVVGGRKVLVSTLERK 834
Db 789 PDLVLDARSDLPTMEVYQVRVLRPADCCSYTISFTTVEIISTRRRAVEATLENRG 848
Qy 835 ENAVNTSLIIFSRNLHLASTLPQRESPIKVECAFS--AHARLCSYGVHPVOTGAKYTF 892
Db 849 ENAVSAVLNIGQSNLQFASLIQDSDNSIECVNEERLHKYCNVSPFFRAKAKYAF 908
Qy 893 LLEFPGSCSLLSQVFGKLTSSPSLEBNGTLQENTQATSAYIYQEBHLFSSBSTLHRY 952
Db 909 RLDEFKSVFLHNLQIHLGAGSDSHRODSTADNTALRLRHLKYEADVLFTRSSLSHF 968
Qy 953 EVHNYGTLPV--GPGEFKTTLRVQNLGCVVSGLIISALLPAVAHGNYFLSLSQVITN 1010
Db 969 EVKANSSLESYDGGPPNVCVFKVQNLGFFPIHGVMMKITPIATRGGRNRLMLRDFPTD 1028
Qy 1011 --NASC-IVQNLTEPPGPVHPELQHTNRNLNGSNTQOVVRCHIGQLAKTEVSGILR 1067
Db 1029 QGNSTCNWGSTYRSTFTE-EDLSHAPQNHNSNDVSIICNL-RLAPQGEFISFYLVG 1086
Qy 1068 LVHNEFFRRAKFKSLTVVSTFELGTEBGSVQLTEASBWSLLEAVQTRPLISLWTLI 1127
Db 1087 NLWLTSLKALKYRSLIKTVNNAALOROFHSPIFREEDPSRQVTFEISKOEDMOVPIWII 1146
Qy 1128 GSVVGLLILLALLVFCMLKGFPAHK-----IPEERK 1160
Db 1147 GSTIGGLLLALLVALLMKLGFFKSAKRKREPGIGLPIKEIK 1188

```

RESULT 14
 AAB30929
 ID AAB30929 standard; Protein; 1188 AA.

XX
 AC AAB30929;
 XX

DT 02-APR-2001 (first entry)
XX Amino acid sequence of a human alpha1 integrin chain.
DE
XX
XX Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;
KM osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
KM wound healing; trauma; rheumatoid arthritis; osteoarthritis;
XX osteoporosis; cartilage damage; bone damage; cartilage.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..22
FT /note="signal peptide"
FT 951..972
FT Region /note="leucine zipper"
FT 1142..1164
FT Domain /note="transmembrane domain"
FT
PN WO200075187-A1.
XX
XX 14-DEC-2000.
XX
XX 31-MAY-2000; 2000WO-SE01135.
XX
XX 03-JUN-1999; 99SE-0002056.
XX
XX (ACT1-) ACTIVE BIOTECH AB.
XX
XX Gullberg D.
PI
DR MPI; 2001-071061/08.
DR N-PSDB; AAC66871.
XX
XX
XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
PT alpha 11 in association with subunit beta; useful for treating muscle
PT dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis -
XX
XX Disclosure: Fig 2a-c; 79pp; English.
XX
XX The present sequence represents a human integrin subunit, designated
CC alpha11. The alpha11 polynucleotide and polypeptide are useful as
CC markers of cell target molecules, such as fibroblasts, muscle cells,
CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
CC They are also used for determining the differential stage of cells
CC during differentiation, development in pathological conditions, in
CC tissue regeneration, in transplantation or in therapeutic and
CC physiological repair of tissues. The pathological conditions involving
CC subunit alpha11 are selected from damage of cells, muscle dystrophy,
CC fibrosis, wound healing, trauma, rheumatoid arthritis, osteoarthritis
CC and osteoporosis, damage of cartilage and bone, and cartilage and bone
CC diseases. The polypeptide is useful for detecting the formation of
CC cartilage during embryonic development, for detecting physiological
CC therapeutic repair of cartilage and muscle, for selection and analysis,
CC or for sorting, isolating or purification of chondrocytes and muscle
CC cells, for detecting regeneration of cartilage or chondrocytes during
CC transplantation of cartilage or chondrocytes during transplantation of
CC cartilage or chondrocytes, respectively, or of muscle or muscle cells
CC during transplantation of muscle or muscle cells, respectively, and for
CC studies of differentiation or chondrocytes or muscle cells.
XX
SQ Sequence 1188 AA;
Query Match 39.4%; Score 2403.5; DB 22; Length 1188;
Best Local Similarity 43.0%; Pred. No. 1.2e-204;
Matches 514; Conservative 212; Mismatches 425; Indels 45; Gaps 16;
QY 1 MELPEFTHLFLPLVFLTGLCSPPNLDENHRLFPGRPEAFGYSVLQHVGGQRMVLVGA 60
DB 1 MDLPRLVAVAMLSLMPGFTDPTFMDTRKRVIPGSRTPAFGTVYQOHISGNKMLVYGA 60
QY 61 PWDGPGDRRGDYRRCVPGAGAHNAPCAKHLGDYQLGSSHPAVNNHMLQMSLLETDDGCG 120

DB 61 PLETNGYOKTGDVYKCPV---IHGNCITKLNLGRVTLNSVNSERKDNMRGLSLATNPKDNS 117
QY 121 FMACAPLMSRACGSSVFSSGICARVDASFOGSLAPTAPORCPTMDVTVLDGNSITYP 180
DB 118 FLACSPFLMSHCGSSYITTGMCNRVNSFRSKYTAAPLQRCQTMVLDGNSITYP 177
QY 181 MSEVOTFLRLYKGLFIDPEQIQVGLVOYGESPVHWSLGFRTKEEVRAAKNLSRREG 240
DB 178 WVEVQHFLINLKRYIGPGQIQVGVDVHVEFLNDRYSKVDVVEAASHIEQKGG 237
QY 241 RETKTAQALMAVACTGFSQSHGRPEARLNVVTVDGSSHDEELPALMKACEARVRY 300
DB 238 TETRTAFGIEFARSAFOK--GGRGAKKAMIVITIDGSHSPDLKVIQSERDNVRY 295
QY 301 GIAVGHYLRORDSSFLREIRTAASPDERFFENVDEAALTIDVALGDRIFGLEGS 360
DB 236 AVAVLGYNRKGINETFLNELKIYASPPDKHFNVTDEALKIYDALGRITSLEGT 355
QY 361 HAENSSFGLEMSQIGFSTHRLKDGILFGMGVAYDMGGSVLMLEGHRLFPERMALEDEF 420
DB 356 N-KNETSFGLEMSQIGFSSHVEDGVLLGAYAYDMNGAVLKETSAKVIPLRESYLKEF 414
QY 421 PPALONHAAYLCYSVSMILKGRRLPLSGAPRFRHREKVAFOLKKGAVRVAOSLOGE 480
DB 415 PEELKNHGAAYLGTVTSVSSROGRVYVAGAPREFNHTKVLFTMHNRSRLTIHQAMEGQ 474
QY 481 QIGSYFGESELCPLDPRDGTDLVLAAPMFLGPONKEGRVYVYLVGQOSLLTLOGTLQ 540
DB 475 QIGSYFGESEITSVLDGQVTDVLLVGAAPMFPN-BGRRGKYYVEL-RQNRFPVNGTLK 532
QY 541 PEPP-QDARFGFAMGALPDLNQDGFADYAVAGAPLEDHGOALYLYHGTQSGVRPAPORI 599
DB 533 DSHSYQNAKRFSSIASVARDNQDSYNDVVGAPLEDNAGAIYIPHGRGSLTKPKQRI 592
QY 600 AAASPHALSTFGKRVQDGLDGDLDVAVAGAGAAIILSSRYVLTSLSEVTPAI 659
DB 533 TASELATLQYFGGCIHQDLNEDGLIDLAVGALGNVILMSRVVQINSLHPEPKI 652
QY 660 SVVQRCRRRQGEAVCLTAAALCFQVTSRTPGMDHQFMRFTASLDEMTAGARAAPDSG 719
DB 653 NIFHRDCKRSGRDATCLAAFLCTFTPIFLAHPQTTTQVIRINATMDERKYPRALDSCG 712
QY 720 QRLSPRLRLSGVNTCEQLHFHVLDSYLRPAVALTVTALDNTKGPVLNCGSPSI 779
DB 713 DRFTRAVILSSGQGLCERINFHVLDYAVKPVFVSVEYSLDEPDH-GPMLDDGMPRTL 771
QY 760 OKTVFSPSDCGPDNECVTVLVQVNMDF-----RSSRK-----APRVV 817
DB 772 RVSVFPMNGCNEDEHCVLDVLDARSDLPTAMEYQQRVLARKPAPODCSAYTLSFDTVTFTI 831
QY 818 RGRKRVAVSTTLERKENAVNTSLISIFSRNLHLASTLPQRESPIKVECAAPS--AAHAR 875
DB 832 ESTRRVAVENTLERGENAYSTVINISQANLQASLQKEDSGSIECNNEERLQKQ 891
QY 876 LCSVGHVFGTAKVTLLEEFSCSSLLSOVFGKLTASDSLSERNGLQENTAOYTSAYI 935
DB 892 VCNVSYPPFRAKAKAVFLDSEFSKSIPLHLEIELAAGSDNSNERDSTKEDVAVLRFHL 951
QY 936 QYEPHLLSSSESTLHRYEVHXYGLP--VGGEPEKTLRLQNLGCVVSVGLIISALLPA 993
DB 952 KYEADVLFTRSSSLSHYEVKLNSSLERVDYGIQPPSCIFRIQNLGLFPIHGMMKITPI 1011
QY 994 VAHGNYFLSLSOVVTN--NASC-IYQVLTPEPGPVPELOHTNRNLNGSTQCOVVC 1050
DB 1012 ATRSGNRLKLKPLDPLTDANTSCINWGSTERYPRPVE-EDLRARPQLNHSNSDVSLINC 1070
QY 1051 HLGOLAKGTEVSGLLRVHNEFFRRAKFKSLTVVSTFELGTBEGSVLQLTASRWSSSL 1110
DB 1071 NI-RLVPOEINFHLGLWMLRSALKYKMKIKMIVNAALORFSPPIFDEEDPSQIE 1129
QY 1111 LEVOTRPLISLWLTLSVVGULLLALLVFLCMLKLGFPFAHKKIPPEEKKEBKE 1166
DB 1130 FEISKQEDMVOVPIWIVGSTLGGILLALLVIALRLKLGFFRSAR---RRRPPGLD 1181

RESULT 15
 AAB25590
 ID AAB25590 standard; protein; 1034 AA.
 XX AAB25590;
 AC AAB25590;
 DT 21-NOV-2000 (first entry)
 XX
 DE Protein encoded by human secreted protein gene #7 clone HOBY69.
 XX
 KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
 KW anticancer; vulnery; antiviral; antibacterial; antifungal;
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
 OS Homo sapiens.
 XX
 PN MO200029435-A1.
 XX
 PD 25-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US25031.
 XX
 PR 28-OCT-1998; 98US-0105971.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JT, Moore PA, Wei Y;
 PI Greene JM;
 DR WPI; 2000-387742/33.
 XX
 PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 PT
 PS Claim 1; Page 678-682; 803pp; English.
 XX
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAB25590-AAB25593 encode the 12 secreted protein sequences given in
 CC AAB25576-A25593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antiarthritic; antirheumatic; dermatological;
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnery;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraproctoloma and myocardiitis;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #7 and protein
 CC sequences are represented in sequences AAB25582 and AAB25582. Secreted
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
 CC AAB25582-A25593 represent genes which are related to the secreted protein
 CC gene7.
 XX
 SQ Sequence 1034 AA;

Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;
 QY 1 MELPVTHLFLPLVLTGLCSFNDENHRLPFPAPAEFGYSLQVGGQRMVLVGA 60
 Db 1 MDLPRLGVVAMALSLMPGFTDTFNNDTKKPRVIPSRSRAFRGTYQOHDISNKKLVVGA 60
 QY 61 EWDGSGRRGDVYRCVPGAHNAPCAKHGLADYOLGNSSHPAVNMHLSILETDGCG 120
 Db 61 PLETNGYOKGTGDVYKCPV---IHGNCITKLNGRVLTSNVSEKDMRKLGLSLATPDKNS 117
 QY 121 FMACAPLMSRACGSSVFSSGICARVDAFQPGSLAPAPQCPPTMDVYIVDGSNSTYP 180
 Db 118 FLACSPLMSSHGSSVYTTGMCSSRVNSNFRSKVTAPALQRCQYMDVIVLDGSNSTYP 177
 QY 181 WSEVOTFRLRLKFLIPEDIOVLVQYGESPVHMSLGPRTXEYVRAKNSREG 240
 Db 178 WEVEHFLINLIKRYIGRGOQYGVVQYGEVYHRLNDRSKYDVEASHLEQKCG 237
 QY 241 RETKTAQAIMVACTBGFQSHGPREARLLVYVVDGESHGDEELPALKACEAGRVTY 300
 Db 238 TETRTAFGIEFARSAFQK--GGRKAKKVMIVITDGESHSDSPDLEKVIQGESRDNVTRY 295
 QY 301 GIANTGHLRRORDSSFLREIRTAASPDERFFNNVDEALTDIVDALGRIGLGS 360
 Db 296 AAVLVGYNNRGINDETFLNEIKYIASDPDDGHFFNVTDKALKDIYDALGRIFSLRG 355
 QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGMVAYMGGSVLMLEGGHRLFPPRMALDEF 420
 Db 356 N-KNETSFGLMSQIGFSSHVYEDVLLGAVGAYMNGAVLKETISAGKVIPLRESYLKEF 414
 QY 421 PPALONHAAYLYGSVSMILRGRLFLSGARFRHRKVIAPOLKKGAVFAVQSLQGE 480
 Db 415 PEELKNHAGAYLYGTVTSVSSRQGRVYVAGAPRFNHTKVLIFTHNNRSLTIHQAMGQ 474
 QY 481 QIGSYFSGELCPDTPDRGTTDVLVAAPMFLGPONKETERVYVVLVQOQSLLTQGTIQ 540
 Db 475 QIGSYFSGEILTSVDIDGGVDVLLVGAAPMFN--BGRBRKAYVEL--RQNFVYNGTLK 532
 QY 541 PEPP--QDARFGPAMGALPDLDNODGPADVAVAGAPLEDGQAGLYLHGTOSGRPPAPRI 599
 Db 533 DSHVQNNARFGSSIASVVDLNDQSYNDVYVAGAPLEDNAGALYIHGGRGSLTKPKORI 592
 QY 600 AAASMPHALSYFGRSVDRLDLDGDDLVAVAGAAIILSSRPVHLTPSLLEVTPQAI 659
 Db 593 TASELATGLQYFGGSIHQGLDLENDGLIDLAVAGALGNVILMSRPVQIINSLHREPSKI 652
 QY 660 SVYQDCCRRCGEAVCLTALCFOVTSRTPGEMHOFYMRFTASLDEWTAGARAAFDQSG 719
 Db 653 NIFHRDCCRSGRDATCLAAFLCFEPIFLAHPHOTTGIRVATWDEKRYTPRAHLDGQ 712
 QY 720 QRLSPRLRLISGVNTCEQLHNVLDTSQDYLBPVALTYTPALDNTTKGPVLENGSPSI 779
 Db 713 DFTFRRAVLLSSGQELCRINHVLDYVAKPVFSVYSLEDDPH--GPMLDGMPITLV 771
 QY 780 QKLVFSSKDCGPDNECVTLVLOVMDI-----RGSRK-----APFVY 817
 Db 772 RVSVPFMWGNEDHCVDIVLDARSDDLPTAMEYQGRVLRKPAQDCSAVTLSTFTVTFII 831
 QY 818 RGRKRVLVSTLTENRKENAVNTSISTFSRNHLASLTPORESPIKYECAAPS--AAR 875
 Db 832 ESTORVAVVEATLENRGNNAVSTVINISQSANLQFASIIQKEDSGSIECVNERLQKQ 891
 QY 876 LCSVGHAPVQCAKXTFLLEPFSSSLISQVFGKLTASSLSLENGTLQENTQATSAVI 935
 Db 892 VCNVSPFPRAKAKAFRLDFEFSSKIFLHHELEIAGSDSNEKSTEDVAPLRFLL 951
 QY 936 QYEPHLLFSSSESTLRYVHVHPTLP--VGPGEKKTLYRQNLQCVVSGIILSALLPA 993
 Db 952 KYEADVLTFRSSSLSHVEVXKNSLSLERDYGIGPESCIFRIONGLFIHGMKITTPI 1011
 QY 994 VAHGNVYLSLSOVITNNASCTIV 1016
 Db 1012 ATRSGNRLLKLRDPLTDEGEHVL 1034

Query Match 36.9%; Score 2254; DB 21; Length 1034;
 Best Local Similarity 44.5%; Pred. No. 2.2e-191;

Wed Jul 16 08:48:31 2003

us-09-647-544-2.open.rag

Page 18

Search completed: July 16, 2003, 07:52:04
Job time : 135.667 secs

118 FLACPLMSHEGSSYYTTCMGSRVNSNFRFSKTVAPALORCQTYMDIVILDGNSIYP 177
181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOGESPVHEMSLQDPFRTKEEVRAAKNLRRRG 240
178 WVEVGHFLINILKRFYIGRQIOVGVQYGEDVHVEHFLNDYRSVADVEAASHIEQGG 237
241 RETKTAQAIWVACTEGFSQSHGRPEARLLVVVTGESHGDEELPAALKACEAGVTRY 300
238 TETRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHSDSDLEKVIQOSERDVTTRY 295
301 GLAVGHYLRQRDPSSFLREITIASDPERFFNVNTPAALTDIVDALGRIIFGLSGS 360
296 AVALVGYNNRGINPFTLEIYIASDDPKHFNVTDEAALKDIDLADGDIIFSLGEG 355
361 HAENESSFGLMSQIGFSTHRLKDGLFGVNGAYDMGGSVLMLEGHRLFPPRMALDEDF 420
356 N-KNETSFGLEMSQTFSSHVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414
421 PPALONHAAVLYGYSVSSMLRGRRLFLSGAPFRHRGKVIAPOLKKDGAVRVAQSGLOGE 480
415 PEELKNHGAAYLGTYTVSVSSRQGRVYVAGAPRFNHTGKVIPLFTMHNNRSLTIHQAMRG 474
481 QIGSYFGESELCPLDPRDGTDTVLVAAPMFLGPQNKETGRVYVYLVGOOSLITLOGTQ 540
475 QIGSYFGESEITVDIDGDTVLVGAAPYFN--EGREGRKVVYEL--RQNRVYNGTLK 532
541 PEPP--QDARFGFAMGALPDINODGFADVAAGAPLEDGHOGALYLYHGTOSGVRPAPORI 599
533 DSHSYQNAFPGSSIASVRDLNODSYNDVVGAGALEDNHGAIIYIFHGFGSLIKTPKORI 592
600 AAASMPHALSYFGRSVDRDLDDGDDLVDAVGAQAAIILSRPIVHLTPSLEVTPOAI 659
593 TASELATGLOYFGCSIHQOLDLNEDEGLIDLAVGALNNAVILMSRPVQOINASHPEPSKI 652
660 SVVQDCCRGRQGAVALCLTALCFQVTSRTPGRMDHOFMRFTASLDEMTAGARAADSGG 719
653 NIFHRDCKRSGRADCLAAFLCTPIFLAPHQOTTVGIRYNATMDERRYPRAHLDEGG 712
720 QRLSPRLRLSVGNVTCEQLHFFHVLDTSDYLRFVALVTFFALDNTKPGVILNEGSPTS1 779
713 DRTNNAVILSSQGEICERINFPVLDIADVKRVTSVEGSLIEDPH--GPMDDGMPITL 771
780 QKLVPSKDCGPNCECVTDILVQNMNDI-----RGRK-----APPV 817
772 RVSVPFMNCGNEDEHCVPLDVLDRASDLPTAMEYCORVLRKPAQDCSAAYTISFDTVFI 831
818 RGRARVLYSTLENKENAYNTSLIIPSRNLHLASLTPQRSPIKVCAPAS--AAHR 875
832 ESTROVAVEATLENGENAYSTVLNISQANLQFASLQKEDSDGSIQVNEERLQKO 891
876 LGSVGHFVQTKAKVTLLEFEFSCSLSQVFGKLTASSDSLERNGLQENTAOQTSAYI 935
892 VCNVSVPFRPRAKAKVAFRLDFEFSKSIPLHLLEIELAAGSDSMERSTEDNVAPLRFL 951
936 QVEPHLLFSSESTLHRYEVHRYGTLF--VGPEPEFTTIRAVONLGYVVSGLIISALLPA 993
952 KYEADVLFRSSSLSHYEVNLNSLERVYDGIQPFICRIQNLGFLPHIGIMKITIPI 1011
994 VAHAGVYFSLSGVITN--NASC--IVQNLTEPRGPVHPELQHNRLNGSTQOQVVR 1049
1012 ATRSGRRLKLKRLDPLDEVANISCNWGSTERYPTVE--EDLRAPOLNHSISDVASTN 1070
1050 CHLGOLAKGTEVSVGLRLVHNEFFPRAKFKSLTVSTFELGTEEGSVQLTEASWSS 1109
1071 CNI--RIVPQOEINFLHGLNMLRSLKALKYKSMKIMVNAALQOHPHSPIFRREDPSROI 1129
QY 1110 LIEVQVTRPILISLWILIGSVLGGILLALVFCIMKGLFPAHKKIPEEKKEEKLE 1166
Db 1130 VVEISKQEDMOWPIWIIIGSTIGLILLALLVLAALKLGFRRSAR----RRREPCHD 1182

Sequence 35, Application US/09836353A
Publication No. US20030129685A1
GENERAL INFORMATION:
APPLICANT: NI et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P1
CURRENT APPLICATION NUMBER: US/09/836.353A
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1998-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 35
LENGTH: 1189
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-35

Query Match 39.8%; Score 2429; DB 9; Length 1189;
Best Local Similarity 43.1%; Pred. No. 1,4e-186;
Matches 516; Conservative 213; Mismatches 422; Indels 46; Gaps 16;

1 MELPFTVHLFLPLVFLVLTGLCSFNDENHPRLLFPGPPEAFEGSYLQHVGGGQRMVLVGA 60
1 MDLPGLVYVAMALSLMPGFTDTFNMDTRKPRVIPSRSRAFCYTYQOHDISGNKVLVYGA 60
61 PMDGGSGRRGDVYRCVPYGAHNAFCAKGLADYQOLGNSSHPAVNMHLGMSLLETGDCG 120
61 PLETNGYQKTGDGVYKCP---IHGNCTKLNLGRVTLVSVSERKDMRGLSLATNPKNXS 117
121 FMACAPLMSRACGSSVFSSGICARVDAFQOGSLAPAPORCPTYMDVIVILDGNSIYP 180
118 FLACPLMSHEGSSYYTTCMGSRVNSNFRFSKTVAPALORCQTYMDIVILDGNSIYP 177
181 WSEVOTFLRLVGLKFLIDPEQIOVGLVOGESPVHEMSLQDPFRTKEEVRAAKNLRRRG 240
178 WVEVGHFLINILKRFYIGRQIOVGVQYGEDVHVEHFLNDYRSVADVEAASHIEQGG 237
241 RETKTAQAIWVACTEGFSQSHGRPEARLLVVVTGESHGDEELPAALKACEAGVTRY 300
238 TETRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHSDSDLEKVIQOSERDVTTRY 295
301 GLAVGHYLRQRDPSSFLREITIASDPERFFNVNTPAALTDIVDALGRIIFGLSGS 360
296 AVALVGYNNRGINPFTLEIYIASDDPKHFNVTDEAALKDIDLADGRIIFSLGEG 355
361 HAENESSFGLMSQIGFSTHRLKDGLFGVNGAYDMGGSVLMLEGHRLFPPRMALDEDF 420
356 N-KNETSFGLEMSQTFSSHVEDGVLLGAVGAYDMNGAVLKETSAGKVIPLRESYLKEF 414
421 PPALONHAAVLYGYSVSSMLRGRRLFLSGAPFRHRGKVIAPOLKKDGAVRVAQSGLOGE 480
415 PEELKNHGAAYLGTYTVSVSSRQGRVYVAGAPRFNHTGKVIPLFTMHNNRSLTIHQAMRG 474
481 QIGSYFGESELCPLDPRDGTDTVLVAAPMFLGPQNKETGRVYVYLVGOOSLITLOGTQ 540
475 QIGSYFGESEITVDIDGDTVLVGAAPYFN--EGREGRKVVYEL--RQNRVYNGTLK 532
541 PEPP--QDARFGFAMGALPDINODGFADVAAGAPLEDGHOGALYLYHGTOSGVRPAPORI 599
533 DSHSYQNAFPGSSIASVRDLNODSYNDVVGAGALEDNHGAIIYIFHGFGSLIKTPKORI 592
600 AAASMPHALSYFGRSVDRDLDDGDDLVDAVGAQAAIILSRPIVHLTPSLEVTPOAI 659
593 TASELATGLOYFGCSIHQOLDLNEDEGLIDLAVGALNNAVILMSRPVQOINASHPEPSKI 652
660 SVVQDCCRGRQGAVALCLTALCFQVTSRTPGRMDHOFMRFTASLDEMTAGARAADSGG 719
653 NIFHRDCKRSGRADCLAAFLCTPIFLAPHQOTTVGIRYNATMDERRYPRAHLDEGG 712

QY 720 QRLSPRLRLSVGNVTCQQLHFFVLDTSYLRPAVLTFTALDNTTKPGPVINEGSPTSI 779
Db 723 DFTNRRAVLSSGQELCERINFPVLDLADYVKKVTSVEYSLEDDPH-GPMLDDGWPFTLL 771
QY 780 QKLVPPSSKCCGPDNECVTDVLYQVNMDI-----RGSRK-----APRVV 817
Db 772 RVSVPFMNGCNDEHCVPLVDLARSDDLPTAMEYQCRVLRKPAQDCSAAYTLSPFTVFI 831
QY 818 RGRRRKVLVSTLENNKENAVNTSLISFRNLHLASLTLPQRESPIKVCAPS--AHAR 875
Db 832 ESTROKVAVEATLENNGENAVSTVLNISQANLQFASLTQKEDSDGSIETCVAEERLQKQ 891
QY 876 LCSVGHVPTQTAKYTFLLFEFSSCSLSQVFGKLTASSDSLERNGTLQENTAOQTSAYI 935
Db 892 VCNVSYPPFRKAKAKAFRLDPEFSKSIPLHLEIELAAGSDSNERNSTEDVAPLRHLL 951
QY 936 QYEPHLLSSSESTLHRYEVHPYGLP--VGQPEBKFTTLRVQNLGCVVSGLIISALLPA 993
Db 952 KYEADVLFTRSSSLSHYEVKLNSSLERYDGIQPFSCIFRIQNLGLFPIHGIMMKITPI 1011
QY 994 VAHGGNYFLSLQVITN--NASC-IVONLTPPGPVHPEELQHTNRLNGSNTQCOVVR 1049
Db 1012 ATRSGNRLKLKDFLTDEVAANTSNIWGSTERYPTVVE-EDLRAPQNLHNSNDVYSIN 1070
QY 1050 CHLGOLAKGTEVSVGLRLVHNEFFPRAKFKSLTVVSTPELGTBEGSVQLTEASRMSHS 1109
Db 1071 CNI-RLVNPQELNPHLGLMLRLSLKALKYKSMKIMVNALQRFHSPIRFEEDPSROI 1129
QY 1110 LLEVQTRPILISLWLIGSVLGLLLALLVFCMLKGFPAHKKIPEEKEEKEKLE 1166
Db 1130 VFEISKQEDMQVPIWIVGSTLGLLLALLVIALMKLGFPSRSAR---RRREPGLD 1182

RESULT 3

US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: P489P2
; CURRENT APPLICATION NUMBER: US/09/984.130
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243.792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836.353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198.407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105.971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43

Query Match 36.9%; Score 2254; DB 9; Length 1034;

Best Local Similarity 44.5%; Pred No. 2.1e-174;

Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;

QY 1 MELPFVTHLFLDPLVLTGLCSPPNLDENHRLFPGPPEAFSGSYLVQHVGGQRMVLVGA 60
Db 1 MDLPRLGLVAVMALSLMPGFTDFEMDTRKPRVIPSRTAFPFYITQOHDISGNKMLVVG 60
QY 61 PMDGSGBRGDGVYKCPVGGANAPCAKHLGDYQLGNSSHPAVNMHIGMSLLETDDGCG 120
Db 61 PLEITNGYOKTGVYKCPV---IHGNTKTLNIGRVTLISNVSERKDMNRIGLSTLATNPKD 117

QY 121 FMACAPLMSRACSSVFSFGICARVDASFQPGSLAPTAPORCPTMDVYIYLDGNSNIYP 180
Db 118 FLACSPLMHSHEGSSYITTCMSRNSNFRBSKYAPLALQRCQITMDIYIYLDGNSNIYP 177
QY 181 WSEVQTFRLRLYKFLIDPEQIYQVLYOGESPVHWSLGFRTKEEVYRAAKNLSRREG 240
Db 178 WEVEQHFPLNLIKRFYIGPQIQVGVYQGEDVHVEFLNDYRSVKDVVEASHIEQRCG 237
QY 241 RETKTAQALMACTGCFQSHGREGPEARLNVYVTDGSSHDGEEPLPAALKACEARVTRY 300
Db 238 TETRTAFGIEFARSBAFK--GGRKAKKMIIVITDGSSHSPDLKVIYQOSERNVTRY 295
QY 301 GIAYGHYLRQRDSSFLREIRTIASPDDEFNFVNDDEALITDVALGRIFGLGGS 360
Db 296 AVAVYGINRKGINETLNEIKYIASPDDEKHFNVNDEANLKIYDNLGRITSLBCT 355
QY 361 HAENESSFGLMSQIGFSTHRLKDGILFGWGAAYDMGGSVLMLEGNHRLFPFRMALEDEF 420
Db 356 N-KNETSPGLEMSQGFSSHVEDGVLLGAVAYDMNGAVLKETSAKVYIPLRESYLFKEF 414
QY 421 PVALONHAAYLYGYSVSMMLKGRRLFLSAGAPRFRHKKVIAFOLKKGAVRVAOSLOCE 480
Db 415 PBEELKNHGAAYLYGTVTSVSSRQGRVYVAGAPRFNHTGVLLFTMHNNRSLTIHQAMRQ 474
QY 481 QIGSYFGESELCPDIDRPGTTDVLVLAAPMFLGPONKTEGRVYVLYVGOQSLTLQGTLO 540
Db 475 QIGSYFGEELISVDDIGGCVTDVLLVGAPMFPN-BGRKRGKYVEL-RQNRFPVNGTLK 532
QY 541 PEPP-QDARFGFAMGALPDNLQDGFADYAVAGAPLEDHGOALYLYHGTOSGVRPPHAPRI 599
Db 533 DSHSQNARFSSSIASVADLNQDSYNDVVGAPLEDNAGAIYIYHGRGSLTKPKQRI 592
QY 600 AAASMPHALSYFGKSVDRDLDDGLVDVAVAGAAIILSSRPYVHLSLEVTPOAI 659
Db 593 TASELATGLOYFGCSIHQDLDBEDGLIDLAVGALGNVVIIMSREVVOINASLHPEPKI 652
QY 660 SVVORDCRRGOEAVCLTAAALCFQVTSRTPGRMDQFMRFTASLDEMTAGARAFDQSG 719
Db 653 NIFHRDCRSGRDAICLAAFLCFTPIFLAPHPQTTTVGIRYNATWDEKRYPPRAHLDBCG 712
QY 720 QRLSPRLRLSVGNVTCQQLHFFVLDTSYLRPAVLTFTALDNTTKPGPVINEGSPTSI 779
Db 723 DFTNRRAVLSSGQELCERINFPVLDLADYVKKVTSVEYSLEDDPH-GPMLDDGWPFTLL 771
QY 780 QKLVPPSSKCCGPDNECVTDVLYQVNMDI-----RGSRK-----APRVV 817
Db 772 RVSVPFMNGCNDEHCVPLVDLARSDDLPTAMEYQCRVLRKPAQDCSAAYTLSPFTVFI 831
QY 818 RGRRRKVLVSTLENNKENAVNTSLISFRNLHLASLTLPQRESPIKVCAPS--AHAR 875
Db 832 ESTROKVAVEATLENNGENAVSTVLNISQANLQFASLTQKEDSDGSIETCVAEERLQKQ 891
QY 876 LCSVGHVPTQTAKYTFLLFEFSSCSLSQVFGKLTASSDSLERNGTLQENTAOQTSAYI 935
Db 892 VCNVSYPPFRKAKAKAFRLDPEFSKSIPLHLEIELAAGSDSNERNSTEDVAPLRHLL 951
QY 936 QYEPHLLSSSESTLHRYEVHPYGLP--VGQPEBKFTTLRVQNLGCVVSGLIISALLPA 993
Db 952 KYEADVLFTRSSSLSHYEVKLNSSLERYDGIQPFSCIFRIQNLGLFPIHGIMMKITPI 1011
QY 994 VAHGGNYFLSLQVITNASCIV 1016
Db 1012 ATRSGNRLKLKDFLTDEGEHVL 1034

RESULT 4

US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins

```
FILE REFERENCE: PF489P1
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 43
LENGTH: 1034
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-43
```

```
Query Match      36.9%; Score 2254; DB 9; Length 1034;
Best Local Similarity 44.5%; Pred. No. 2,1e-174;
Matches 464; Conservative 187; Mismatches 356; Indels 36; Gaps 11;
```

```
QY 1 MELPFTHLFLPLVFLGLCSFPLNDBHRLFPPEPEAFGYSVLQHVGGGQRMVLVGA 60
   |||
DB 1 MDLPRLVVAWALSLMPGFTDTFPMOTRRKRVIPGSRTAFEGYVQOHISGKMLVGA 60
   |||
QY 61 PMDPSGDRGDVYRCVPGAHNAPCAKGLGDYQGNSSHPVNMHLGMSLLETGDDG 120
   |||
DB 61 PLENTGYQKTGDYKCPV---IHGNTKLNLGKVTLSNVSERDNRLGLSLATNKDMS 117
   |||
QY 121 PNAACPLMSRACSSVVFSSGICARVDASFOPOGSLAPTORCPTVDVIVILDSNSIYP 180
   |||
DB 118 FLACSLMSHECCSSYTTGMCSSRVANSNFRFSKTVAPALORCQTDMDIYILDSGNSIYP 177
   |||
QY 181 WSVQVFLRLVGLKFLIDPEQIOVGIVQGESPVHMSLGDPTREKVRAAONLSRREG 240
   |||
DB 178 WVEVQFLNLKKEFYIGQOIQGVGVQGVVHFLNDKSVKDVVEASHLEORGG 237
   |||
QY 241 RETKTQALINVACTEESQSHGSRPEARLLVVVTGESHDEGLPALKACAGAVTRY 300
   |||
DB 238 TERRIAFGIEFASEAFQK--GGRKAKKMYITTGESHDSDLKVIQGSERDVTY 295
   |||
QY 301 GLAVLGHYLRORDPSSFLREIRTIASDPDERFFENVTDAAITDVALGDRIFGLEGS 360
   |||
DB 296 AVALGLYNNRGINPEFLNEIKYIASDDPKHFFVNTDEALKDIVDALGDRIFGLEGT 355
   |||
QY 361 HANESSFGLEMSQIGFSTHRLKDGILFGVAGVDMGSGVLMLEGGHRLFPPMALDEDF 420
   |||
DB 356 N-KNETSPGLEMSQIGFSSHVEDGVLLGAVGAYDMNGAVLAKETSAGKVIPLRESTLKEF 414
   |||
QY 421 PPLQNHAAVLYGSVSMILRGRLFLSGAPRRFRGKVIAFOLKKDGAVRVAQSLQGE 480
   |||
DB 415 PBEKLNKHAVALGTVIVSVSSRGGRVYVAGAPRRFNTGKVIILFTMNNNSLTIHQMRQ 474
   |||
QY 481 QIGSYGSELCPILDRDGTDTVLLVAAPWFLGPNKTEGRRVYVILVGOOSLITLQGLTQ 540
   |||
DB 475 QIGSYGSEITSDIDGDVTLVVGAPWYFN--EGREGRKVVYEL--RQNFVYNGTLK 532
   |||
QY 541 PEPP--ODARFPGMGLPDLNODGFADVAGALEGHCAGALVYCTOSGVRPHRQRI 599
   |||
DB 533 DSHSYONARFGSSIAVRDLNODSVADVAGALEHNAHALYIFHGFSGSLIKTPKORI 592
   |||
QY 600 AAASMPHALSYFGRSVYDGLDLDDGLDVVAVAGOGAAILLSRPVHNLTPSLLEVPORI 659
   |||
DB 593 TASELATGQYFEGCSIHGOLDLNEDELIDAVGALGNAYILMSRPVQVQINASHFPPSKI 652
   |||
QY 660 SVVQDRCRRGOEAVLTALCFQVTSRTPGRMWHOFYWRFTASLDEMTAGAPAAFGSG 719
   |||
DB 653 NIFHRKCSKSGRDTAFLAFLCFTPIFLAPHQTTVIGIRYNAIMDEKTYTPRAHIDEG 712
   |||
QY 720 QRLSPRLSLVGNVVCGLHFFVLDTSDYLARVALTVTFALDNTTKRPVNLGSGPTSI 779
   |||
DB 713 DRFTBAVALLSSGOELCEIRINFLVDTADYVVKPVTFSVEVSLDDPDH--GPMLLDDGMPPTL 771
   |||
```

```
QY 780 OKLVFESKDCGPDNECVTDVLQVNMDI-----RGSRK-----APFVY 817
   |||
DB 772 RVSVPFMGNCNDHECHVDIVLARSDLPTAMEYQGVRLKRAQDCSAVTLSTFDTTFVFI 831
   |||
QY 818 RGRKRLVSTLTENRKENAVNTSLSTFSRNLHLASTLPQRESPIKYECAAPS--AHAR 875
   |||
DB 832 ESTROVAVEATLENRGENAVSTVINISQANLQFASIIQKXDSGSIQCVNEERLQK 891
   |||
QY 876 LCSVGHVPGTCARAKTFLLEPEFSSLSQVFGKLTASSDSLEBNGLTOENTQATSAVI 935
   |||
DB 892 VCNVSTPFRKAKAVAFRUDFEFSKIFLHLIEILAGSDSNERDSTKEDNVALRPHL 951
   |||
QY 936 QYEPHLFSSSESTLHRYEVEHPYGLP--VGQPEFKTLRVQNLCCYVSGLIISALLPA 993
   |||
DB 952 KYEADVLFTRSSSLSHYEUKNSSLERYDGLGPPFSCIFRIQNLGLPFIHGMKITPI 1011
   |||
QY 994 VAHGNYFLSLSQVITNNASCIV 1016
   |||
DB 1012 ATRSGNRLKJLDFLTDGEHVL 1034
   |||
```

RESULT 5

```
US-09-984-130-103
; Sequence 103, Application US/09984130
; Publication No. US2003005231A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-103
```

```
Query Match      30.9%; Score 1887; DB 9; Length 1151;
Best Local Similarity 36.3%; Pred. No. 1,9e-144;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;
```

```
QY 23 ENLDEHRLPLFPPEPEAFGYSVLQHVGGGQRMVLVGA PMDPSGDRGDVYRCVGAH 82
   |||
DB 1 FNVDAKNMTEGSPJEDHFGYTVQGYENEKMWVILGSLPLVGOPKRRGADVYKCPVGGE 60
   |||
QY 83 NARCAKHLGDYQGNSSHPAV-----NMHLGMSLLETGDDGGFMAKPLMSRACSSVF 137
   |||
DB 61 SLPCVKLDP---VNTSIPNTEYKEMWTFG--STLVNPNMGGLACGLVAVRGGHLHY 115
   |||
QY 138 SSGICARVDASFOPOGSLAPTORCPTVMDVIVILDSNSIYPMSVOTFLRLVGLKFI 197
   |||
DB 116 TTGICSDVSPTFOVYVNSIAP--VQESCTOLDIYIVILDSGNSIYPMDSVTAFLNDLLKMDI 174
   |||
QY 198 DPEQIOVGLVOYGSSPVHMSLGDFTREYEVRAAKNLSRREGRETKTAQAIMACTGEF 257
   |||
DB 175 GKQIOVQIVQYGENVTHEFNLNKYSSTEEVLVAAKKIVQGGGROTMALGTDAKBAF 234
   |||
QY 258 SOSHGCRPEARLLVVVTGESHSHOGEELPALKACEARVTVRGYAVLGHYLRORDSS 317
   |||
DB 235 TEARGARGVKKVMVIVTDGSHDNHRLKAVIQDCEDENIQRFSTAILGYSYNGNLSTEK 294
   |||
```

```

QY 318 FLEIRTTIASDPERFEFENVITDEALTDIYDALGDRIFGLESGHAEENESSFGLENSQIGF 377
DB 295 FVEBEIKSIASEPTKEKFFNVSDDELAVITVKTGRIIFALEATADQASAFEMENSQIGF 354
QY 378 STHRLKDGILFGWVGAYDMGGSVLMLEGGHRLFPFPMMALEDEFPPLALONHAAYLGYSVS 437
DB 355 SAHYSQDMWMLGAVGAYDMNGVVMQKASQIIPNTTNNVSTKKNBEPLASLYLGTVNS 414
QY 438 MLRGGRRLFLSGAPFRHRGKVIAFOLKKDGAVRVAQSLQEOIGSYFSGELCPLDTR 497
DB 415 ATASGDVLYIAGQPRNHTGVYIYRM-EDGNIKILOTLSEBOIGSYFGSILTTTIDK 473
QY 498 DDTTDLVLAAPMFLGPONKETGRVYVYLVGOOSLTLTGTOF----- 541
DB 474 DNTDILLVGA PMWMTKEKEGQKVVYAL-NQTRFEYQMSLEPIKOTCCSRQHSCTT 532
QY 542 ---EPPODARFGFAMGALPDNLQDGFADVAVGAFLDEHGQALYLYHGTQSVRPHAPOR 598
DB 533 ENKNEPCGARFGTALAAVKDMLDGFNDIVIGAPLEDHGA VYIYHSGKTIKREYQOR 592
QY 539 IAAAGMPHALSYFGHSVDRDLDDGDDLVDAVGAQGAAILSSRPIYHLTPSLEVTPOA 658
DB 593 IPSGDDGKTLKFFGSGIHGEMDNLGDLTVITIGLGAALFWSRDUVAVVKYTNMFEPRK 652
QY 659 ISVORDCRRGOEAVCLTAALCFQVTSRTGPRMDHOFYMRFTASLDEWTAGARAAPDGS 718
DB 653 VNIQKNCHMEGKETVCINATVCFEYKLSKEDTIEADLOYRVTLDSLRQISRFSGT 712
QY 719 GQRLSPRLRLSVGNVTEQOLHFVLDTSYLRPALVYTFALDNTTK--GPVINEGSP 776
DB 713 QORRVOR--NITVRKSECTKHSFYMLDKGDFODSVRIITLDF--NLTPENGVPVLDLSLP 767
QY 777 TSIOKLVPFSKDCGPNDECVTDLVLOVNMIDIRGSRKAPFVVGRRKVLVSTLENRKEN 836
DB 768 NSVHEIYIPRAKOCCKEKCISDLSLHV---ATTEKDLIYRSQNDKRVSLIVKNTDS 823
QY 837 AYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPSAAHRLCSVGHVPFOTGAKVTFLEEF 896
DB 824 AYNTRTIYHVSPLVFSGI---EAIQKQSC--ESNHNITCKVGPFLRGEWTFKILF 877
QY 897 ESSCSLSLQVFGKLTASDSELRNCTLOENATQTSAYIQYEPHLLFSESESLHRYEVP 956
DB 878 QNTSYLMEVNTIYISATISDSEPEPTLSDNVNISIIPKVEYGLQFYSASEYHISIDA 937
QY 957 YGTLF-----VGPGEFTTLRVONLGCYVSGLIISALLPAVAHGN---YFLSLQV 1007
DB 938 NETVEVINSTEDIGNIINIFYLIRKSGSFPMPELKLISFRMNSNGFPUYPTGLSS- 996
QY 1008 ITNNAACIVQNLTEP-----DGPVHPEBLQHTNRNLGNSNTQOQVVRCHLQOLAKTEV 1061
DB 997 -SENANCRPHIFEDPFSINGKMTTSTDHLKRGITLDCNTCKPATITCNLTS-SDISGV 1054
QY 1062 SVGLRLVHNEFFRAKFKSLIVVSTFELGTBEGSVLQUTEASRNSSESLLEVQTRPL 1120
DB 1055 NVSL--ILMKPTFISYFSSLMLTRGELRSENAS-LVSSSNQKELAIQISKGLPGR 1111
QY 1121 ISLWILIGSVLGLLLALLLVFCMLWLGFEFAHKKIPEEKREE 1163
DB 1112 VFLWVILLSAFAGLLMLLMLLILALMKITGF---KRLDKKMER 1151

```

RESULT 6
US-09-836-353A-103
; Sequence 103, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:

; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19

```

; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-103

Query Match      30.9%; Score 1887; DB 9; Length 1151;
Best Local Similarity 36.3%; Pred. No. 1.9e-144;
Matches 429; Conservative 226; Mismatches 454; Indels 74; Gaps 22;

QY 23 FVLDHHRRLFPPEARFEGSYVLOHVGCGRMLVGA PMDPSDRDRGDVYRCVGAH 82
DB 1 FNVDRKNSMTFSGPVEDMFGYTVQOYENEGKMWLIGSPLVQQPKRTGDVYKCPVGRGE 60
QY 83 NAPCAKGLIGDYQLGNSHPAV-----NMHLGMSLETDGDDGFMACAPLMSRACGSSVF 137
DB 61 SLPCVKLDP-----VNTSIPNTEVKEKMTFG-STLVNPNNGFLACGPLIAYRCGLHY 115
QY 138 SSGICARYDASFQPGSLAPTAORCPYMDVVIYLDGNSIYPMSEVOTFLRLVYKLEFI 197
DB 116 TTGICSDVSPFQVNSIAP--VQECSTQDLDIYIVLDGNSIYPMDSVTAFLNDLKRMDI 174
QY 198 DPEQLOVGLVOYGESEPVHMSLGDRTKEEYVRAKANUSRREGRTKTAQAIMACTGCF 257
DB 175 GPKQOYQIVQYGEVNTHEFNLNKYSSTEEVLAANKKIVQORGRQTMALGTDTRAKAF 234
QY 258 SOSHGREGAARLLVNVVDGESHDELPALAKACEAGVTRVYGA VGHYLRQDRPSS 317
DB 235 TEARARRGVKKWYIVTDGESHDRKLKVIQDCEDENIQFSAIIGSYNRKGLSTBK 294
QY 318 FLEIRTTIASDPERFEFENVITDEALTDIYDALGDRIFGLESGHAEENESSFGLENSQIGF 377
DB 295 FVEBEIKSIASEPTKEKFFNVSDDELAVITVKTGRIIFALEATADQASAFEMENSQIGF 354
QY 378 STHRLKDGILFGWVGAYDMGGSVLMLEGGHRLFPFPMMALEDEFPPLALONHAAYLGYSVS 437
DB 355 SAHYSQDMWMLGAVGAYDMNGVVMQKASQIIPNTTNNVSTKKNBEPLASLYLGTVNS 414
QY 438 MLRGGRRLFLSGAPFRHRGKVIAFOLKKDGAVRVAQSLQEOIGSYFSGELCPLDTR 497
DB 415 ATASGDVLYIAGQPRNHTGVYIYRM-EDGNIKILOTLSEBOIGSYFGSILTTTIDK 473
QY 498 DDTTDLVLAAPMFLGPONKETGRVYVYLVGOOSLTLTGTOF----- 541
DB 474 DNTDILLVGA PMWMTKEKEGQKVVYAL-NQTRFEYQMSLEPIKOTCCSRQHSCTT 532
QY 542 ---EPPODARFGFAMGALPDNLQDGFADVAVGAFLDEHGQALYLYHGTQSVRPHAPOR 598
DB 533 ENKNEPCGARFGTALAAVKDMLDGFNDIVIGAPLEDHGA VYIYHSGKTIKREYQOR 592
QY 539 IAAAGMPHALSYFGHSVDRDLDDGDDLVDAVGAQGAAILSSRPIYHLTPSLEVTPOA 658
DB 593 IPSGDDGKTLKFFGSGIHGEMDNLGDLTVITIGLGAALFWSRDUVAVVKYTNMFEPRK 652
QY 659 ISVORDCRRGOEAVCLTAALCFQVTSRTGPRMDHOFYMRFTASLDEWTAGARAAPDGS 718
DB 653 VNIQKNCHMEGKETVCINATVCFEYKLSKEDTIEADLOYRVTLDSLRQISRFSGT 712
QY 719 GQRLSPRLRLSVGNVTEQOLHFVLDTSYLRPALVYTFALDNTTK--GPVINEGSP 776
DB 713 QORRVOR--NITVRKSECTKHSFYMLDKGDFODSVRIITLDF--NLTPENGVPVLDLSLP 767
QY 777 TSIOKLVPFSKDCGPNDECVTDLVLOVNMIDIRGSRKAPFVVGRRKVLVSTLENRKEN 836
DB 768 NSVHEIYIPRAKOCCKEKCISDLSLHV---ATTEKDLIYRSQNDKRVSLIVKNTDS 823
QY 837 AYNTSLSIIFSRNLHLASLTPQRESPIKVECAAPSAAHRLCSVGHVPFOTGAKVTFLEEF 896

```

```

Db      824 AYNRTTTHYSNVLVPSGI-----EAIQKDSG--ESNHNITCKVGFPLIRGEMVTFKLLF 877
QY      897 EFSGSSLLSQVFGKLTASDSIERNGTLOENTAQTSAYI QYEPHLLFSSSESTLHRYEVP 956
Db      878 QFNITSYLMENVITLYLATSDESEEPETLSDNVNINISIPVKEVGLQFYSASAEYHISIAA 937
QY      957 YGTLR-----VGPRPEFTKTLRVQNLGVVSGLLISALLPVAHAGN---YFSLTSQV 1007
Db      938 NETVEPVINSTEDIDONEINIPYLKSSGFPMPEIKLISIPFNMTSNQYPVILYPTGLSS 996
QY      1008 ITNNASCIYONLTP-----PGPVEPHBELOHTRLNGSNTOCOVVRCHLGQLAKTEV 1061
Db      997 -SENANCPHIFEDFPSINSCKMKTSTSDHLKRGILLDCNTCKFATITCNULTS-SDISQV 1054
QY      1062 SVGLIRLVHNEFFRRAPAKSKLTVTSFELGTBEGSVLOUTEASRWSBLEVYQT-RPIL 1120
Db      1055 NVSLT-IIMKPTFIKSYSSLLLTIRGELRSENAS-LVLSASNQRELAIOISKGLPGR 1111
QY      1121 ISIMTLIGSVLGGILLALLVPCVCKLKLGPFAHKRIPEEKRE 1163
Db      1112 VPLWVITLSAPAGLLMLMLITLAIWKIGF--KAPLKKKHEK 1151

```

```

RESULT 7
US-10-160-354-2
; Sequence 2, Application US/10160354
; Publication No. US20030013107A1
; GENERAL INFORMATION:
; APPLICANT: Chu, Peter
; APPLICANT: Li, Congfen
; APPLICANT: Liao, X' Charlene
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Alpha 2 Integrin: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-001110US
; CURRENT APPLICATION NUMBER: US/10/160,354
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/296,819
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1181
; TYPE: PRM
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: wild type alpha 2 integrin, CD49b
; US-10-160-354-2

```

Query Match	28.8%	Score 1756	DB 9	Length 1181
Best Local Similarity	34.3%	Pred. No. 9, 2e-134		
Matches	418	Conservative 219	Mismatches 455	Indels 126
			Gaps	31
QY	11	LPVLVLTG-----GSPFNIDENHRLPFGPPPEAEPQSYLVHGVGQRMVLGAPMD	63	
Db	11	LPLLVLTLVLSGGILNCCCLAYVNGLEPEAKFISGPPSSEQFPAVQOEINPKGNMLLVGSPFS	70	
QY	64	GPSGDRGRDGYRCPVGGANHAPCARGHL-GDYVLNSSHPAVNMILGMSLLETDDGGFM	122	
Db	71	GFPENRMDDYKCPV-DLSTACEKLTNIQTSTINENVEMKTKNMSLGLILRNMTGGFL	129	
QY	123	ACAPLMSACGSSSVSSSGICARVDSAFOPOGSLAFTAOECPTMYDVVLYLDGSSNIYPMWS	182	
Db	130	TGCPFLMAQCCNQYVTTVCSDISDFOLSAFSPATOPCEFLIDVVVVVCCESNSIYPMD	189	
QY	183	EVQTELRRLVYGLFTFDEPQIVGLVQYESPVHSESLGDFRIKEEVNPAKULSRERE	242	
Db	190	AVKNFLERFVGGLDIGFTKTYGGLIQYANNRPVENVNLTMYKTKKEMIYATISQTSQYGGDL	249	
QY	243	TKTQAQINVACTEGSSQSHGGRPEARLLVYVTTDOESHDEELPAALKAACEAGRYTRYGI	302	
Db	250	TNTFCALIDYARKRYAASAGGRSRSTKYMVVVTTDSEHDSGSLTKAVIDQCHNDILIRGI	309	

```

QY 303 AVLGHYLRQRODSSFLREITITSDDPERFENVTPBEALTDIVLADGBIFGLEGSHA 362
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 AVLQGLNNALDITKNLIKELIKALISIPERYFVNSDAALEAKRGTGBOIFISIEG-V 368
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 363 ENESSFGLMEMQIGFST--HRLKAGILFGWGAADVMDGWSVLW-LEGHRLPPPMALADE 419
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 369 QCGDNFQWEMSQVGFSA DYSSQNDILMLGAVGACMSGTIYQKXSHGLIFP-----KQA 423
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 420 FPPALQ--NHAAYIGYSSVSMILRGRRFLISGAPRFRHGRKVIATFOLKQDAVRVAOSL 477
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 424 FDQLLOBRNHSYGYGVAA--ISTGESTHFVAGAPRANYTQIYLVSYNENGINITVIOAH 482
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 478 QGEOIGSFGEICLPDTRDCTDVLVYAPMLGLQNKRTGVVVVY-----LVGOOSL 532
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 483 RGDIGISFGSVLCSVDVDDKDTIDVLLVGAAPMWSDLKKEEGVVYFTTIKGLIQGQOF 542
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 533 LTLQGLQPEPPODARFGFAGALPDLINODAFADAVAGAPLEDHOGALYLYHGTSQSVR 592
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 543 --LEG--PEBIEINTRGSALALSDINDMGFNDVIGSPLEONSGAVITYNHGQIGIR 597
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 593 PHPAQRIAAA--SMPHALSYFGRSVDRDLDDGDLVDVAVGAQAAIILSSRPVILHLP 650
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 598 TKYSQKILGSDGARFRRSHQYGFRLSDGDLNGDSITDVSIGAFGVQVQLWQSIAVAI 657
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 651 SLEVTPOAIVSYQDCCRRCRGAEAVCLTALALCFQYTSRTPGWBHQFVYMRFTASIDEMWAG 710
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 EASFTEPKITILVNNKAAQ-----IILKLCFSAKFR-PTKONNOVAIVNIITLDA--- 704
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 711 ARAAFDGGGQRLSPRRLL-----RLSYGVN-----TCEOLHFVHLDTSDVLRPAVALTVT 758
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 705 ----DGSASVYTRGLEKKNNECCLOKMWVYVNOAQSCPEHIIYIOESDVVNSDLRLVD 759
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 759 FALDNTTKPG--PVLANEGSPTSIOKLVPFSDGCPDNECVTDVLVQVNMIDIRSGRKAPV 816
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 760 ISLEN---PGTSPALAEVSEFAKVFSPFHKDCGEDGICISDLVLVDR-QIPAQEOBFI 815
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 817 VRGRKRVKLVSTLENNKENAYNNTSLIIFSRNHLASLUTQRRSPV--KVCE-APSA 872
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 816 VSNONKRLTFEVTILKNNKRESAYVGIADVEENLFFAFS---LPADGTEVTCQVAVASQ 871
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 873 HARLCSVGHVPFOTGAKVTFLEPFESCSSILSOVFGLTATSSDSLERNGLJOENTAOQS 932
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 872 KSVACDVGYPLAKEOQVTFITINDPULQNLQNASISFQALSSQENKRA--DNLVNLK 929
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 933 AYIOYEPHLLFSSESTLHRYEVHBYGLP-----VGPPEFKTLRAVONIGCVVSGLI 986
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 930 IPLLVDABIHUTRSTINIFEYBISDGNVPISVHSFEDVGPXFIISLKV-TTGSVPVSMAT 988
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 987 ISALLPVAHOGNYFLLSOYITNNACIYONLTERP-----GPPVHEELQHTNRL 1038
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 989 VTIHPIQVTEKBNPLMYLTGVQTDKADISCNADINPLKIGOTSSSVSFKSENRRHTEL 1048
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1039 NGSNTOCOVNRCHLGOLAKTEVSVGLRLVCHNEFFPRAKSKSLTVVSTFELGEEGVL 1098
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1049 NCRFASCSNTCWLKDVMKCEYVNVYTRIMNGTFASSSTFQIYQULAAAEINTYNEBIY 1108
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1099 QUTEARSMSESILEVQTRPILISLWT-----LIGSVLGGLLLLALLVFCWL 1145
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1109 -----VIDNVTIPLIMKPKDEKAEVPGVIGISIIAGIILLALLVALIWL 1154
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1146 KLGFPAHK-----KIPBE 1158
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1155 KLGFKKRYEKMTNPNDE 1172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 8
 US-10-125-540-313
 ; Sequence 313, Application US/10125540
 ; Publication No. US2003005875A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

```

? FILE REFERENCE: PT214C1
? CURRENT APPLICATION NUMBER: US/10/125,540
? CURRENT FILING DATE: 2002-04-19
? Prior Application removed - See file Wrapper or Paim
? NUMBER OF SEQ ID NOS: 646
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 313
? LENGTH: 707
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-10-125-540-313

```

Query Match	28.3%	Score 1727.5	-DB 9	Length 707
Best Local Similarity	49.8%	Pred. No. 6.6e-132		
Matches 340	Conservative 123	Mismatches 109	Indels 9	Gaps 6

RESULT 9
 US-09-764-870-313
 : Sequence 313, Application US/09764870
 : Patent No. US20020042386A1
 : GENERAL INFORMATION:
 : APPLICANT: Rosen et al.
 : TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 : FILE REFERENCE: P1714

```

; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-764-870-313

```

Query Match	28.3%	Score 1727.5	DB 10	length 707
Best Local Similarity	49.8%	Pred. No. 8.6e-132		
Matches 340	Conservative 125	Mismatches 209	Indels 9	Gaps 6

RESULT 10
US-09-691-943-2
: Sequence 2, Application US/09891943
: Publication No. US20030077278A1
: GENERAL INFORMATION:
: APPLICANT: Galatin, Michael W.
: APPLICANT: Van der Vlieten, Monica
: TITLE OF INVENTION: No. US030030077278A1el Human 2
: FILE REFERENCE: 27866/35004

```

; CURRENT APPLICATION NUMBER: US/09/891,943
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-891-943-2

Query Match      18.5%; Score 1131.5; DB 9; Length 1161;
Best Local Similarity 29.4%; Pred. No. 6,6e-83;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRPLFPGRPEAFEGSYVLQVGGGQRMMLVGAAPWDGSGDRGDRVRCPPVGAH 82
DB 17 FNLDVEEPTIFQ-EDAGFGGOSVVOF--GGSR-LVVGAPLEVVAAVAAQTGRLYDC---AAA 69
QY 83 NAPCAGHIGDYOLGNSHPAVNMHIGMSLLETGDDGGFMACAPLMSRACGSSVFSGIC 142
DB 70 TGMCOPIPL-----HIREPAVNMISGLTLASTNGSRLLACGPTLHRVCGENSYSKSGC 123
QY 143 ARVDASFPOGSLAFTAQRCP-TYMDVVIYLDGNSI--YPMSEVQTLRLVGLKFLIDP 199
DB 124 LLLGSRWEIITQVDPATPECPHOEMDIVFLIDSGSIDQNDFMQMGFQAVWGQ--FEG 181
QY 200 EQQVQVLVQGESPEVHEMSIGDRTEKEVYRAKNSREGRRTKQAQIMACTGEGFQ 259
DB 182 TDTLFLMCOYSMLKIHFTFTQSPSQSLVDPIVLQKGL-TFTATGILTVVTLDFHH 240
QY 260 SHGGRPEARLAVVVTDEGSH--DGEELPALKACBAARVTRVGIAVLGYLRQRPDSF 318
DB 241 KNGARSAKKILVITDQKYKDPLEYSDVIPAQEKAGIIRVAGV--GHAF---QGPFA- 295
QY 319 LREIRTIASDPPERFFPNVTDEALTDIYDALGDRIFGLEGSHAENESSFGELEMSQIGFS 378
DB 296 ROELNITISSAPQODHFVKYDNFALGSIQKOLEKIYAVEGTQSRASSFQHEMSQEGFS 355
QY 379 THRLKGIIFGMYGADVWGVSVMLEGGHRLFFPRMALDEFPALQONIA-----AY 430
DB 356 TALTMGDLFLGAVGFSW-----SGAFLYPPNMS-----PTFINMSQENVMDRDSY 402
QY 431 LGYSVSMILRGGRRLFLSGAPRRHGRKVIAF-QLKKGDAVVAOSLOGEOIGSYFGSE 489
DB 403 LGSTLALMKQGNVL--GAPRYCHTKAVITQVSRQW--RKKAEVGTGTGIFGAS 459
QY 490 LCPDLTDRTDVTDLVLAAPMLGPNKETGRVYVVLV--GQGSLLTLQGLQRPDPD- 546
DB 460 LCSVDVDSGSDTLIIGAPHY--EQTRGGQVSVCLPRGQVQOCDAVLRGEGGHW 517
QY 547 ARFGFMALPDLNOCGFADVANGARLEDBGHQALHYNG--TOSGRPHRAPQRIAAAMP 605
DB 518 GRRGALTVLGVNNEKLIIDVALGARGEBENRGAVVLFPGASGSIISPHSQSIASSQS 577
QY 606 HALSYGRSVGDLDDGDLVDVAGAGAAAILLSSRPVHLTPPLEVTTPQALISVQGD 665
DB 578 PRLOYTGQALSGGDDITDGBMLAVGARGQVILLLSLPLVKVGMAMRSPEVEAAVVR 637
QY 666 C-----RRRGEAVCLTAALCFQYTSRTPGRMHQFYWRFTASLDENTAGARAAFDGS 718
DB 638 CMEKPSALEAGATVCLTIQ-----KSLDQGLDQSSVRFDLALDPELTSRAIFNET 692
QY 719 GQRLSRRLRLISGANTCEQLHNVLD--TSYLRPALVTPALDNTTKRG-----PVAN 772
DB 693 KQPTLTRRRTLIG-THCETLKLPLDCEVEDVVSPIILHNSLVAREPIPSQNTLRPVL 751
```

```

QY 773 EGSPTSIQKLVFPSKDCGPDNECVTDLVLOVMMDIRGSRKAPFVVGRRKVLVSTLLEN 832
DB 752 VSSQDLFTASLPFEKNCODGICEBD--LGVTLSGSIQT---LVGSSLELNAVITYWN 806
QY 833 RRENAVNTSLSTIFSRNHLASLTPQRESP---IKVEC-AAPSAHARL---CSVGHV 883
DB 807 AEDSYGVTVASLYYPAGLSHRVSGAQKQPHOSALRLACETVPTDEGRSSRCVNHPI 866
QY 884 FOTGAKVTFLEEFEESSCSLLSQVFGKLTASSDSLERNGTLOENATQTSAYIQYEPHLE 943
DB 867 FHEGSGNGTIVTFVDYSYATLADRM-LMPASGSSNNKRASSKATFQLEPVKIVVTVI 925
QY 944 SSESTLHREYVAPYCTLFPVGPPEFRTTLRYQNLCCYVVGSLIISALLPAVAGNGYPLS 1003
DB 926 SNGEESTKY--FNPAISDEKQKWEKHEHRYVANNLSQRDLA-LSINFWPVLVINGVAVMDV 982
QY 1004 LSQVITNNASCTVQNLTEPPGPVPAPELOHTNR--LNGSNTQCVVYRCHGLQAKTE 1060
DB 983 VNEASQSLPCVSEK-----KPPQHSDFLTQISRSPMLDCSIADCLQRCDDVPSPVOEE 1037
QY 1061 VSVGLRLVHNEFFRRAKFKSLTVSTFELGTEEGSVQLTEASRMSSELE-VVQTRPI 1119
DB 1038 LDFTLKGNLSFGWVETLQKVLVSVVAITFDTSVYQOLPQQAFAFRQEMEMVLEDBV 1097
QY 1120 LISLILIGSVLGLLALLLVFCLMKLGF--AHKKIPEEKREE 1163
DB 1098 YNAIPIWSSVGLALLLALITATLYKLGFFRHHKEMLEDPED 1142

RESULT 11
US-09-350-259-2
; Sequence 2, Application US/09350259
; Patent No. US20020062008A1
; GENERAL INFORMATION:
; APPLICANT: Galatin, Michael W.
; TITLE OF INVENTION: No. US20020062008A1 Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/350,259
; EARLIER FILING DATE: 1999-07-08
; EARLIER APPLICATION NUMBER: 09/193,043
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-350-259-2

Query Match      18.5%; Score 1131.5; DB 10; Length 1161;
Best Local Similarity 29.4%; Pred. No. 6,6e-83;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRPLFPGRPEAFEGSYVLQVGGGQRMMLVGAAPWDGSGDRGDRVRCPPVGAH 82
DB 17 FNLDVEEPTIFQ-EDAGFGGOSVVOF--GGSR-LVVGAPLEVVAAVAAQTGRLYDC---AAA 69
QY 83 NAPCAGHIGDYOLGNSHPAVNMHIGMSLLETGDDGGFMACAPLMSRACGSSVFSGIC 142
DB 70 TGMCOPIPL-----HIREPAVNMISGLTLASTNGSRLLACGPTLHRVCGENSYSKSGC 123
QY 143 ARVDASFPOGSLAFTAQRCP-TYMDVVIYLDGNSI--YPMSEVQTLRLVGLKFLIDP 199
```



```

Db      124 LLLGSRWEIIQIVPDPATPECPHOEMDIVLIDGSSGIDQNDNFNMOKGFIQVAMGQ--FEQ 181
Qy      200 EOIQVGLVOYGESPVHEMSLDGPFRTKEEVRAAKNLSREGETKTAQAINVACTEGESQ 259
Db      182 TDTLFLMQLYNSLNKHFHTFOFRITSPSOOSLVDPIVQJLGL-TFTATGILVVVQLPFH 240
Qy      260 SHGGRPEARLLVVTGDESH--DGEELPALKACAGATRYGIAVLGHYLRORDPSF 318
Db      241 KNGARSAKKILIVITDQKYPKPLEYSVDIIPAERKAGIIRYAIGV-GHAF--QGPTA- 295
Qy      319 LREIRTIASDPDERFFENVTDAAALTDIDALGDRIFGLESHAHNESSFGLIEMSGIFS 378
Db      296 ROELNTISSAPPODHVFKVDNFPALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 355
Qy      379 THRLLKGLIFGNVAGAYDMGSLVLMLEGHRLFPPRMALDEFPALONHA-----AY 430
Db      356 TALTMGDLFLGAVGFSW-----SGAFIYPRNMS-----PTFINNSQENVMDRSDY 402
Qy      431 LGYSVSMULRGGRRLFLSGAPFRHRKGVIAF-QLKDGAVRAVQOSIQEIGSYFSGE 489
Db      403 LGYSTELALMKGVQNLVL-GAPRYOHTGKAVIFTQVSQW--RKALEVGTQIGSYFGAS 459
Qy      490 LCPIDTRDGTVDVLLVAPMFLGPONKETGRVYVVLV--GQOSLLTLOGTQIPEPPD- 546
Db      460 LCSVDVDSGSTDLLILIGAPHYY--EOTRGQGVSCPLPRGRQVQWQCDAVLRGEQHPW 517
Qy      547 ARFGFAMGALPDINODGFADVAVGAPLEDEHOGALTYLHG-TOSGVRPHPAORIAAAS 605
Db      518 GRFGAALTIVLGDVNEKILIDVAIGAPGEENRGAVALPFGASBSGISPSHSQRIASSQS 577
Qy      606 HALSYFGRSVDGRLLDGGDLVDVAVGAQGAAILLSRPVHLTPSLEVTPOAISVQD 665
Db      578 PRLQYFGALSGGODLTQGLMDLAVGARGOYLLLRSLPVLKGVAMRFSPEVAKAVYR 637
Qy      666 C-----RRROEAVCLTALALCFQVTSRTPGRMDFQFMRTASLIDENITAGARA 718
Db      638 CMEKPSALEADGATVCLTIQ-----KSSLDQIGDIQSSRPDLADPRLTSLRALFNET 692
Qy      719 GORLSRRLRLSVGNVTCQLHFNHVD-TSDYLRPVALTVTFALDNTTKPG-----PVAN 772
Db      693 KNPITLTKRTLGLG-IHCETILKLLPDCYEDVVSPIILHNSLVAREPIPSQNLRPVLA 751
Qy      773 EGSPTSIOQLVPFSKDCGPDNECVTDVLQVNMDIRGSRKAPFVVRGGRKYLVTSTLEN 832
Db      752 VGSODLFTASLPFEKRCQGDGLCEGD--LGVTLSFGGLQI--LTVGSSLEINVIYTVMN 806
Qy      833 KRENAYNTSLSTIFSNLHLASLTPORESP---ITVECAAPSAAHARL-----CSYGHV 883
Db      807 AGEDSGYIVVSLYYPAGLSHRVVSQAQKQPHQSALRLACETVPTEDGSRSSRCVNHPI 866
Qy      884 FQTGAKVTFLEPEFSCSLISOVFGKLTASSDSLERNGLTOENTQOTSAYIOYEPHLLF 943
Db      867 FHEGSGTIVTFDVSYKATLGDRLM-LMPSASSENKKAASSKATQLELPKXAVAYYMI 925
Qy      944 SSESTLHRYEHPYGTLPVGPGEFEKTLRLVONLGCYVVSGLIISALLPVAHAGNYFLS 1003
Db      926 SHQESTKY--FNPAISDEKMKKEAHRVANNLSGRDLA-ISINWVVLINGVAVMVPV 982
Qy      1004 LSOVITNNASCIYONLTERPPRPVARELQHTNR--LNGSNTQCVARCHLQOLAKGTE 1060
Db      983 VNEASQSILPCVSR-----KPPHSDPLTQISRSPLMDCSIADICQFCDVPSPFQDE 1037
Qy      1061 VSVGLLRLVHNEFPRAKFKSLTVVSTFELGTEEGSVLQLTASRWSESLLE-VYOTRI 1119
Db      1038 LDFTLKGNISFGWRETTLOKAVLVVSAEITPDTSYSQLPGOEAMRWOMENVALEDEV 1097
Qy      1120 LISLMTLIGSVLGLLLALLVFCIMKLGF-AHKKIPEEKREE 1163
Db      1098 YNAIPIIMGSSVCALLLLALLITATLYKLGPFKRHYEMLEDKRED 1142

```

RESULT 12
US-09-891-943-99

```

; Sequence 99, Application US/09891943
; Publication No. US20030077278A1
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. US20030077278A1el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/891,943
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-891-943-99

Query Match      18.4%; Score 1126; DB 9; Length 1161;
Best Local Similarity 29.5%; Pred. No.1,9e-82;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

Qy      23 ENLDEHPRLLPFGPEAEFGYSVLQHVGGQRMVLVGAIPMDGPSGRRGDYRCPVGAH 82
Db      18 FLNDEEPTLIFQ-EDAGFGQSVOF--GGR-LTVGAPLELVANQGRLYDC---AAA 70
Qy      83 NARCAKGLIDYQLGNSHPVNMHLMGMSLLETDDSGFPMCAPLMSRACSSVSQSGIC 142
Db      71 TGMCCPTPL-----HIREAVNMSLGLTLASTNGSLACGPTLHRVCGENSTSKSC 124
Qy      143 ARVDASFOPOGSLAFTAORCP-TYMDVIVILDGNSNI--YPMSEVQTEFLRLVLGKLFIDP 199
Db      125 LLLGSRWEIIQIVPDPATPECPHOEMDIVLIDGSSGIDQNDNFNMOKGFIQVAMGQ--FEQ 182
Qy      200 EOIQVGLVOYGESPVHEMSLDGPFRTKEEVRAAKNLSREGETKTAQAINVACTEGESQ 259
Db      183 TDTLFLMQLYNSLNKHFHTFOFRITSPSOOSLVDPIVQJLGL-TFTATGILVVVQLPFH 241
Qy      260 SHGGRPEARLLVVTGDESH--DGEELPALKACAGATRYGIAVLGHYLRORDPSF 318
Db      242 KNGARSAKKILIVITDQKYPKPLEYSVDIIPAERKAGIIRYAIGV-GHAF--QGPTA- 296
Qy      319 LREIRTIASDPDERFFENVTDAAALTDIDALGDRIFGLESHAHNESSFGLIEMSGIFS 378
Db      297 ROELNTISSAPPODHVFKVDNFPALGSIQKQLOEKIYAVEGTQSRASSSFQHEMSQEGFS 356
Qy      379 THRLLKGLIFGNVAGAYDMGSLVLMLEGHRLFPPRMALDEFPALONHA-----AY 430
Db      356 TALTMGDLFLGAVGFSW-----SGAFIYPRNMS-----PTFINNSQENVMDRSDY 403
Qy      431 LGYSVSMULRGGRRLFLSGAPFRHRKGVIAF-QLKDGAVRAVQOSIQEIGSYFSGE 489
Db      404 LGYSTELALMKGVQNLVL-GAPRYOHTGKAVIFTQVSQW--RKALEVGTQIGSYFGAS 460
Qy      490 LCPIDTRDGTVDVLLVAPMFLGPONKETGRVYVVLV--LVGQOSLLTLOGTQIPEPP 544
Db      461 LCSVDVDSGSTDLLILIGAPHYY--EOTRGQGVSCPLPRGRQVQWQCDAVLRGE-QGHP- 516
Qy      545 QARFGFAMGALPDINODGFADVAVGAPLEDEHOGALTYLHG-TOSGVRPHPAORIAAAS 603
Db      517 WGRFGAALTIVLGDVNEKILIDVAIGAPGEENRGAVALPFGASBSGISPSHSQRIASSQ 575
Qy      604 MPHASTYFGRSVDGRLLDGGDLVDVAVGAQGAAILLSRPVHLTPSLEVTPOAISVQ 663
Db      576 LSPRLQYFGALSGGODLTQGLMDLAVGARGOYLLLRSLPVLKGVAMRFSPEVAKAV 635

```

QY 664 RDC-----RRGOEAVCLTAAICFOVTSRTGPRMDHOFYMRFTASLDEMTAGARAAD 716
 Db 636 YRCWEKPSALBAGDAVTCCTIO-----KSLDQDGIQSSVRFDALDPGRILTSRAIFN 690
 QY 717 GSGQRLSPRLRLSVGNVTCQJLHFVLD-TSDYLAPVALTVTFALDNTTKPG-----PY 770
 Db 691 ETKNPFLTRKRTGLG-ICETLKLPLDCEVDVSPILHLNFSLVREPIPSQNLRPV 749
 QY 771 LNEGSPTSIOKLVFESKDCGPDNECVTDVLQVNMIDGSRKAPFVVRGGRVLVSTTL 830
 Db 750 LAVGSDPLFTASLPFEKNCGQDGLCEGD--LGVTLSFSGLOT--LTVGSSLELNIYIV 804
 QY 831 ENRKENAVNTSLISITFSRNLHLASLTPORES-----IKVEC-AAPSAHARL-----CSYGH 881
 Db 805 WNAEGESYGTAVSLYYPAGLSHRVSGAQKQPHQSMALRLACETVPTDEGLRSRCSVNH 864
 QY 882 PVFQTAQVTLFLEFEFSCSSLSQVFGKLTASDSLERNGTLQENTAGTSAYIOYEPHL 941
 Db 865 PTFHESNGTFIVTFPVSYKATLGDRM-LMRASASENNKASSSKATFOLELPVKYAVYT 923
 QY 942 LFSESEGLHRYEVHPYGTLPVGPPEFKTTLRVONIGCYVSGLIISALLPAVAHGNYF 1001
 Db 924 MISROESTKY--FNATSEKMKKEAHRVNNLSQDLA-ISINFVPUVLNGVAVW 980
 QY 1002 LLSQVITNNASCIYQNLTEPPGPVHPBELQHTNR--LNGSNTQCVVRCHLGQLAKG 1058
 Db 981 DVWMEAPQSJPCVSR-----KPOHSDPLTQISRSPMLDCSIADCLGPRCDVPFSVQ 1035
 QY 1059 TEVSUGLRLVHNEFFRAKFKSLTVVSTFELGTEGVSVOQLTEASRMSESLLE-VVQTR 1117
 Db 1036 BELDFTLKGNSLFGWRETLOKKVLVVAELITFTDTSYSQLPQGEAFMRQMEVLEED 1095
 QY 1118 PILISIMILIGSVLGLLALLVFLCMKLGFF-AHKKIPEEKREE 1163
 Db 1096 EYVNAIPITMGSSVGAALLLALITATLYKLGFKKRYKEMLEDKPED 1142

RESULT 13
 US-09-350-259-99
 ; Sequence 99, Application US/09350259
 ; Patent No. US20020062008A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gallatin, Michael W.
 ; APPLICANT: Van der Vieren, Monica
 ; TITLE OF INVENTION: No. US20020062008A1el Human 2
 ; FILE REFERENCE: 27866/35004
 ; CURRENT APPLICATION NUMBER: US/09/350, 259
 ; CURRENT FILING DATE: 1999-07-08
 ; EARLIER APPLICATION NUMBER: 09/193, 043
 ; EARLIER FILING DATE: 1998-11-16
 ; EARLIER APPLICATION NUMBER: 08/173, 497
 ; EARLIER FILING DATE: 1993-12-23
 ; EARLIER APPLICATION NUMBER: 08/286, 889
 ; EARLIER FILING DATE: 1994-08-05
 ; EARLIER APPLICATION NUMBER: 08/362, 652
 ; EARLIER FILING DATE: 1994-12-21
 ; EARLIER APPLICATION NUMBER: 08/943, 363
 ; EARLIER FILING DATE: 1997-10-03
 ; NUMBER OF SEQ ID NOS: 114
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 99
 ; LENGTH: 1161
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-350-259-99

Query Match 18.4%; Score 1126; DB 10; Length 1161;
 Best Local Similarity 29.5%; Pzed. No. 1.9e-82;
 Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

QY 23 ENLDENHRLRFRPRRAEGSYVLQHVGGQRMVLGAPWDSGSGRRDYYRCPVGGAH 82
 Db 18 FNDLVEEPTIFQ-EDAGFGQSVQF--GGRS-LVVGAPLEVVAAQGTRLYDC--AAA 70

QY 83 NAPCAKHLGDYOLGNSGHPAVNNHLMGSLLETTDDGGFMACAPLMSBACSSVPSGIC 142
 Db 71 TGMQCPILP-----HIRPEAVNMSLGLTLASTNGSLNACGPTLHHVCGNSYSKSC 124
 QY 143 ARVDASFPQGSIAFTAQRC- TYWDVVIYVUDGSNTI--YPMSEVQTLRLVGLFTDP 199
 Db 125 LILGSRWEIITVPDATERCEPHQEMDIVFLIDGSSGIQDNFQNMKGCFVQAMGO--FEG 182
 QY 200 BQIQVGLQYGGSPPHWESLSDGFRTEKEVNAAKNLRSREGETTAQAIWACREGESQ 259
 Db 183 TDTLPAQVSNLKIHFTHFTQFTSPSQOSLVDPITVQKGL-ITFATGILTVTQTLHH 241
 QY 260 SHGGRPEARLLVNTDSESH-DGEBELPAALACACAGVTRYGIAVLGHYLRORPSSP 318
 Db 242 KMGARKSAKILIVITDQKYKDPLEYSQVLPQAKGAIIRALGV-GHAF--QGPIA- 296
 QY 319 LREITIASDPERFEFVNTDEALTDIVDALGRIFGLEGSNAENESSFGLENSQIGFS 378
 Db 297 ROELNTISSAPQDHFVFDVNFALGSIQKLOEKIYAVEGTQSRASSSFOHEMSQEGFS 356
 QY 379 THRLEDGLFQWVGAIVMGSGSVLMEGGHRLPRPMALADERPPALQNH-----AY 430
 Db 357 TALTYMDGLFLGAVGSFSW-----SGAPFLYPPNMS-----PTFIMNSQENVDMRDSY 403
 QY 431 LGYSVSMILRGGRFLFSGAPRFRHRGKVIAF-QLKXGAVRVVAQSLQGEQIGSYFSGE 489
 Db 404 LGYSTELALMKGVQVQLV-GAPRYOHTGKAVIFITQVSNQW--RKAEVLTGQIGSYFGAS 460
 QY 490 LCPDLTDDGTTDVLLVAAPMFLGPONKETGRVYVY-----LVGQOSLLTLOGTLOPEBP 544
 Db 461 LCSVDVDSGTDLLIGAPPHY--EOTRGQVSCPLPRGRVQCCAVLRGE-QGAP- 516
 QY 545 QDARFPMGALPDNODGFADVAVAGAPLEBQHGALVLYHG-TQSGVRPPHQAORIAAS 603
 Db 517 -MGRFGAALTVLGDVNEBKLDIVALGAPGEOBNKCAVTLFHGASGSGISPSHSQRIASSQ 575
 QY 604 MPHALSYFGRSYDGLDDGDDLVDAVAGAGAAIILSSRPVHLTPSLLEVTPQAISVYQ 663
 Db 576 LSPRLQYFGQALSGQDITDQGLMDLAVGARQVLLNLSLPLVKKVGMRRSPVEVAKAV 635
 QY 664 RDC-----RRGOEAVCLTAAICFOVTSRTGPRMDHOFYMRFTASLDEMTAGARAAD 716
 Db 636 YRCWEKPSALBAGDAVTCCTIO-----KSLDQDGIQSSVRFDALDPGRILTSRAIFN 690
 QY 717 GSGQRLSPRLRLSVGNVTCQJLHFVLD-TSDYLAPVALTVTFALDNTTKPG-----PY 770
 Db 691 ETKNPFLTRKRTGLG-ICETLKLPLDCEVDVSPILHLNFSLVREPIPSQNLRPV 749
 QY 771 LNEGSPTSIOKLVFESKDCGPDNECVTDVLQVNMIDGSRKAPFVVRGGRVLVSTTL 830
 Db 750 LAVGSDPLFTASLPFEKNCGQDGLCEGD--LGVTLSFSGLOT--LTVGSSLELNIYIV 804
 QY 831 ENRKENAVNTSLISITFSRNLHLASLTPORES-----IKVEC-AAPSAHARL-----CSYGH 881
 Db 805 WNAEGESYGTAVSLYYPAGLSHRVSGAQKQPHQSMALRLACETVPTDEGLRSRCSVNH 864
 QY 882 PVFQTAQVTLFLEFEFSCSSLSQVFGKLTASDSLERNGTLQENTAGTSAYIOYEPHL 941
 Db 865 PTFHESNGTFIVTFPVSYKATLGDRM-LMRASASENNKASSSKATFOLELPVKYAVYT 923
 QY 942 LFSESEGLHRYEVHPYGTLPVGPPEFKTTLRVONIGCYVSGLIISALLPAVAHGNYF 1001
 Db 924 MISROESTKY--FNATSEKMKKEAHRVNNLSQDLA-ISINFVPUVLNGVAVW 980
 QY 1002 LLSQVITNNASCIYQNLTEPPGPVHPBELQHTNR--LNGSNTQCVVRCHLGQLAKG 1058
 Db 981 DVWMEAPQSJPCVSR-----KPOHSDPLTQISRSPMLDCSIADCLGPRCDVPFSVQ 1035
 QY 1059 TEVSUGLRLVHNEFFRAKFKSLTVVSTFELGTEGVSVOQLTEASRMSESLLE-VVQTR 1117
 Db 1036 BELDFTLKGNSLFGWRETLOKKVLVVAELITFTDTSYSQLPQGEAFMRQMEVLEED 1095

Query Match 18.4%; Score 1124.5; DB 10; Length 1161;
 Best Local Similarity 30.1%; Pred. No. 2.5e-22;
 Matches 358; Conservative 207; Mismatches 486; Indels 139; Gaps 45;

QY 20 CSEFNDIHHPLPPEPEAEFGYSVLQHVGGQRMVLGAPMDGSGDRGDVRCPLYG 79
 DB 17 CHGSNLDVEEPIVF-REDAASFQOTVVOF--GGSR-LVWGAPLEAAVAVQGTGLXDC--- 69
 QY 80 GANNAACAGHLDYOLGNSHRAVMHIGMSLLETDGGGFGACALPMSRACGSSVFS 139
 DB 70 ---AP-ATGMCOPIYL-RSPLAVNMSJGLSLVTNTNNAQLACGPTQACVKNMYAK 123
 QY 140 GICARVASFQPGSLAFTAQRCPY-MDVVILDGNSI--YPMSEVOTFLRLVKGLE 196
 DB 124 GSCLLGSSLQFIQAVPASMPECPRQEMDIAFLIDSGSINGQDFQMDPVYALMGE-F 182
 QY 197 IDPEQIQVGLVOYGESPVHEWSLGDFTTEEVVRAAKNLSRREGRETQAIIMVACTG 256
 DB 183 ASTSTL-PSLMOYSNLKTHFTPEFKNLIDPSLVDPIVOLQGL-TYTATGIRVMEEL 240
 QY 257 FSGSHGRPEEARLLVVVTDGESH-DGELPALKACEAGRVTRVGIIVLGHYLRQRDP 315
 DB 241 FHSKNGSRKSAKKLLVITDQKRDPLEYSDVIPADAKAIIIRVAVGDAF---QEP 296
 QY 316 SSFLREIRTIASDPERFFFNVTDEALTDIVDALGDRIFGLESHAENESSFGLMSQI 375
 DB 297 TA-LKELNTIGSAPQGDHVPKVGNFALNSIQKLOEKIFALIGTQSRSSSFQHEMSQE 355
 QY 376 GSGTHRLKQGLFGMVGAYDMGGSVLMLEGHRLPPP-----RMALDEFPPLQNH 428
 DB 356 GFSSALTSQPVLGAVGSSFSW-----SGGAFLYPPNTRPTFINMSOEN-----VDMRD 403
 QY 429 AYLYGVSSMLRGGRFLFSGAPRRHKGVIAPQLKKQGAVRVAQSLQGEQISFGS 488
 DB 404 SYLGISTAAAFKGVNSLIL-GAPRHQHTGKVIF-TQEARHMRPKSEVRGTQISYFGA 461
 QY 489 ELCPDLTDRDGTDLVLAAPMELGPONKRTGRVYVLYV-GQOSLTLTQTLQPEBPD- 546
 DB 462 SLCSVDVDRDGSXDLVLIGAPHY--EOTRGQGVSVFVPVGRVGRWQCEATLHGEGHPW 519
 QY 547 ARFGFAMGALPDINOGFADVAVGAPLEBDHOGALYHNG-TOSGVRRPRAQRIAAAMP 605
 DB 520 GRFGVALTVLGDVNGNLDVAIGARGEESRGAVYIFGASRLIEIMPSPSORVTGSOLS 579
 QY 606 HALSYGRSVDGRLDDGDVLVAVAGAACATILSSRPVHLTPSLVTPQALSVVQRD 665
 DB 580 LRLQVFGSLSGGQDLTQDGLVDLAVGAGHVLILRSPLIKVELSIRAPMEVAKAVYQ 639
 QY 666 CRRR-----GQEAVALTAALCFQVTSRTPGRMDH-QFYMRFTASLDEWTAGARAAPDG 717
 DB 640 CWERTPTVLEAGATVCLT-----VHKSPDLIGNVQGSVRYDLALDPGLISRALFDE 693
 QY 718 SCGRSLSPRLRLSVGNVTCBQLHFVLD-TSDYLRVALTVTFAL-DNTTKP---GPVLN 772
 DB 694 TNACTLTGRKTLGLGD-HCEIVKLLPDCVEDAVSPILRLNPSLVRDSASPRNLHPVLA 752
 QY 773 EGSPTSIGLIVPSKDCGPDNECVTDVLQVNMIDIGSRKAPFVVGGRKVLVSTLLEN 832
 DB 753 VGSQDHITASLPEPKCKOELLCEBDLGISFNFS-----GLOVLVGGSPBELTVTYVWN 807
 QY 833 KRENAVNTSLSLIFSRNLHLASLTPQRES---PIKVEC-AAPSAHARL-----CSVGHPV 884
 DB 808 EGEDSYGTLVKFYYPAGLSYRRTVGTQGHQVPLRLACEAPRAQEDLRSSSCSINHPIF 867
 QY 885 QTGAKYTFLLPEFSCSSLSLSQVFGKLTASDSLENGTLQENTA-QTSAVYQY----- 937
 DB 868 REGAKTTFMITPDVSYKAFLDRL--LLRAKASSENNKPDINKTAFQLELPVXYTYVTLI 925
 QY 938 -----EPHLFPSS-----ESTLHRYEYHPYGTLPVGPPEFKTLRLVQNLGCVVSGL 985
 DB 926 SROEDSTNHVNFSSSHGGRQEAHRYRYVNNLSPL-----KLAVRVNFVVPVLLNGV 977
 QY 986 II-SALLPAVAGNYFLSLQVITNNASCIVONLTPPGRPVHPELOHTNRLNGSNTQ 1044

DB 978 AWMVDTLSSPAQG-----VSCVSO--MKEPONPDDFLTQIQRSSVLDCSIAD 1021
 QY 1045 CQVACHIGQALAKTEVSVGLRLRVHNEFFRRAKSLTVSTFELGTEEGSVLOL--TE 1102
 DB 1022 CLHFRCDIPLSLDIODELDFILRGNLSPGWSQTLQEKVLVSEALITDTSVYSQLPQE 1081
 QY 1103 A---SRMSSESLLEVQTRPILISLWILIGSVLGGLLALVFCMLKLG 1149
 DB 1082 AFLRAQVETTLFEYVVEPI-----FLVAGSSVGGLLALITVLYXKLG 1127

Search completed: July 16, 2003, 08:12:49
 Job time : 89.5295 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 45.9578 Seconds
(without alignments)
747.132 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 6106

Sequence: 1 MELPVTHTLPLVPLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1131.5	18.5	1161	1	US-08-173-497-2
2	1131.5	18.5	1161	1	US-08-286-889-2
3	1131.5	18.5	1161	1	US-08-485-618-2
4	1131.5	18.5	1161	1	US-08-362-652-2
5	1131.5	18.5	1161	2	US-08-605-672-2
6	1131.5	18.5	1161	2	US-08-482-293A-2
7	1131.5	18.5	1161	2	US-08-943-363-2
8	1131.5	18.5	1161	4	US-09-193-043-2
9	1131.5	18.5	1161	4	US-09-688-307A-2
10	1126	18.4	1161	2	US-08-485-618-99
11	1126	18.4	1161	2	US-08-605-672-99
12	1126	18.4	1161	2	US-08-482-293A-99
13	1126	18.4	1161	2	US-08-943-363-99
14	1126	18.4	1161	4	US-09-193-043-99
15	1126	18.4	1161	4	US-09-688-307A-99
16	1124.5	18.4	1161	4	US-09-193-043-55
17	1124.5	18.4	1161	4	US-09-688-307A-55
18	1118.5	18.3	1161	1	US-08-485-618-55
19	1118.5	18.3	1161	1	US-08-362-652-55
20	1118.5	18.3	1161	2	US-08-605-672-55
21	1118.5	18.3	1161	2	US-08-482-293A-55
22	1118.5	18.3	1161	2	US-08-943-363-55
23	1109	18.2	1161	1	US-08-485-618-53
24	1109	18.2	1161	1	US-08-362-652-53
25	1109	18.2	1161	2	US-08-605-672-53
26	1109	18.2	1161	2	US-08-482-293A-53
27	1109	18.2	1161	2	US-08-943-363-53

28	1109	18.2	1161	4	US-09-193-043-53	Sequence 53, Appl
29	1109	18.2	1161	4	US-09-688-307A-53	Sequence 53, Appl
30	1108.5	18.2	1151	1	US-08-286-889-37	Sequence 37, Appl
31	1108.5	18.2	1151	1	US-08-485-618-37	Sequence 37, Appl
32	1108.5	18.2	1151	1	US-08-362-652-37	Sequence 37, Appl
33	1108.5	18.2	1151	2	US-08-605-672-37	Sequence 37, Appl
34	1108.5	18.2	1151	2	US-08-482-293A-37	Sequence 37, Appl
35	1108.5	18.2	1151	2	US-08-943-363-37	Sequence 37, Appl
36	1108.5	18.2	1151	4	US-09-193-043-37	Sequence 37, Appl
37	1108.5	18.2	1151	4	US-09-688-307A-37	Sequence 37, Appl
38	1105	18.1	1170	2	US-08-789-078-2	Sequence 2, Appl
39	1105	18.1	1170	2	US-08-752-633-2	Sequence 2, Appl
40	1105	18.1	1170	5	PCT-US95-04686-2	Sequence 2, Appl
41	1103	18.1	1155	1	US-08-286-889-46	Sequence 46, Appl
42	1103	18.1	1155	1	US-08-485-618-46	Sequence 46, Appl
43	1103	18.1	1155	1	US-08-362-652-46	Sequence 46, Appl
44	1103	18.1	1155	2	US-08-605-672-46	Sequence 46, Appl
45	1103	18.1	1155	2	US-08-482-293A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-08-173-497-2
; Sequence 2, Application US/08173497

; Patent No. 5437958

; GENERAL INFORMATION:

; APPLICANT: Gallatin, W. Michael

; APPLICANT: Van Der Vliet, Monica

; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha

; NUMBER OF SEQUENCES: 29

; CORRESPONDENCE ADDRESS:

; ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 233 S. Wacker Drive, 6300 Sears Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606-6402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/173,497

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5437958and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 27866/31363

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-474-6300

; TELEFAX: 312-474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO. 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1161 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-173-497-2

Query Match 18.5%; Score 1131.5; DB 1; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDHHRRLPGPEAFGYSVLCHVGGGRRMLVGNPMGPGSGDRGVDYRCVGAH 82
DB 17 FNLDVEPTIFQ-EDAGGFGSVQF--GGSR-LVGGAPLEVVANQGRLYDC---AAA 69

```

QY 83NAPCAKGNHGDVQLGSSHPAVNMHNGMSLLETDGSGFMAKCPWMSRACGSSVPSFGIC 142
Db 70TGMCPBPL-----HIREPAVNMISGLTILMASTNGSRLLDCPPLTHRRCGEISYSKSG 123
QY 143ARVDAFPCQGSIAPIAPACRP--TYMDVVIVLDSNSI--YPMSEVOTFLRLTGKFLPID 199
Db 124LLGSGWEIIQVYPDATPECPHDEMIVFLIDSSGSIIDQNDFNQMGKFIQAVWGO--FEG 181
QY 200EOIQVGLVOYGESPVHEWSLGDFTRIKEVEVRAAKNLSRREGRETKAQAIMWACTEFSQ 259
Db 182TDTLFLMOMYSNLKIHFTFOTFRITSPQOSLVDPIVOLKGL--TFPATGILVVTLOLFHN 240
QY 260SHGGRPEARLILVVTYDGESEH--DGEELPAALAKKCEAGRTYRGIYAVLGHYLRQRDPSSF 318
Db 241KNGARKSAKKILIVITDGQYKDPLEYSVITPOAEKAGIIRIYAGV--GHAF--QGPYA- 295
QY 319LREIRTIASDPDERFFNFYNTDEALTDIYDALGDRIFGLEGSHAENESSFGLMESQIGFS 378
Db 296ROBLNTISAPPODHFYKNDNFALGSIQKQLOEKIYAVEGTOSRASSSFQHEMSGEGFS 355
QY 379THRLKDGLIFGVAVGADWGGSVLWMEGHRLEPPRNALDEFPALQNH-----AY 430
Db 356TALTMDGLFLGAVGSSFSW-----SGCAFLYPPNMS-----PTFINMSQENVDMDXY 402
QY 431IGYSVSMMLRGGRFLFISGAPFRFRKGVIAF--QKXGAGVAVVAOSLOGEOIGSYFSGE 489
Db 403IGYSTELAMKGVQNLVL--GAPYQHTGAIVLFTYVSROW--RKKAELVGTQIGSYFSGAS 459
QY 490LCPLTDRDGTDTVLLVAPMFGLPQNKETGRVYVLV--GQOSLTLTLOGTLQDPBPPOD- 546
Db 460LCSVDVDSGSDTLILIGAPHY--EOTRGGOVSCEPLRGQRVQMOCCAVLRGEGHFW 517
QY 547AEFGFAMGALPDLNODGFADVAVGADLEDHOGALYLYHG--TOSGVRPHPAQRIIAASMP 605
Db 518GRFGAALTVLGVDNEBKLDIVAIIGAPGQENRGAVYLFHGAASBSGISPSHQIASSQS 577
QY 606HALSTFGRSVDRDLDDGDDVLDVAVGAGAAIILSSRPVYALHTELEVTPOAISVQGD 665
Db 578PRLQYFGQALSGODTLTQGLMDLAVGAGQVYLLSLPVLKGVAMRFSPEVAAYVR 637
QY 666C-----RRRQGEAVCLTALACFOVTSRTPGRWDRQFMRFPASLDEMTAGABAFDGS 718
Db 638CMEEKPSALAEADATVCLITQ-----KSLIDLQGLDIQSSVRFPLADPGRILSRALFNET 692
QY 719GORLSPRRLSLVGNVTCESQLFHVL--TSDYLRLPVALVTVPALDNTTKPG-----PVIN 772
Db 693KNPILTRKRTLTGIG--IHCETLKLLPLRDCVEDVVSPIILHNFSLVNEPIPSQNLRPVLA 751
QY 773ESSPFSIQCLVFPKDCQCPDNECDVDVLVQVMDINDGSRKAPVRVNGRRKVLVSTLLEN 832
Db 752VQSODLFTASLPEFEKKCCQGDGLCEGD--LGVTLSPSGLOT--LTVGSSLELNVITYWN 806
QY 833RKENAYNTSLSIIFSNNHLASITTPQRESR---IWEVC--AAPSAAVL-----CSVGHV 883
Db 807AGEDBYGTAVSLYYPAGLSHRKYSQAKOPHOSALALACEYPTDEBGLRSSRCSYNHPI 866
QY 884FOTGAKVTFLLEFEFSCSSILSQVFEKLTASSDLSLERNGTLOENTQAOTSAYIOYEBHLF 943
Db 867FHEGSNGFTVFPDVSYKATLQDRM--LMAASASSENKKAASSKATQLELPPVYAYYTM 925
QY 944SSESTLHRYEHPYGTLPVPGCBPEFTTLRVONLGYCYVSGLISALLPAVAHAGNYFIS 1000
Db 926SRQESTKY--FNPAITSDEKKMKKEHRYRVNNLSORDLA--ISINFWVULLANGVAVMDV 982
QY 1004LSQVITNNASCIVONLTERPRGPVHREELQHNTR--LNGSNMQQOVVACHLQOLAKGE 1066
Db 983VNEAPSQSLPCVSESR-----KPPQHSDFLTQISRSMLDCLSIADCLQCFCDVPSFVQEE 103
QY 1061VSVGLRLVHNEFFPRAKFSLTIVVSTFELGTEGSEVLDTASRWSSELE--VVOYTRPI 1111
Db 1038LDFTLKGNLSFCFWAVETTLQKVLVYVASVAIITDTSYSSDLPQGEAMRQOMENVLEDEY 109
QY 1120LISMLILGSLVGLLLALLLVFCIMKLGAF--AHKKIPEEKREE 1163

```

[illegible]

```

296 ROELNTISSAPPODHVFKVDNFAALGSIQKOLEKIVAVEGTQSRASSSFQHEMSQEGFS 355
379 THRDKGILFGMGAVYDMGSSVLMLEGHRLPPRMALEDEPPALONHA-----AY 430
356 TALTDGFLGAVGSSW-----SGAFLYPPNMS-----PTFIMSOENVMRDSY 402
431 LGYSVSMWLLRGGRFLFSGAPFRHGRKVIAP-OLKKGAVRVAOSLOGEOIGSYGSGE 489
403 LGYSTELAMKGVONLVL-GAPRYOHTGKAVIFTQVSRQW--RKAAYVTGTQIGSYFGAS 459
490 LCPPLDTRDGTDLVLAAPMFLGPNKGTGRVYVVLV--GQOSLLTLOGTLOPEPPO- 546
460 LGSVVDSDSGSTDLILIGAPHY--EOTRGQGVSVCPPLRGORVQWQCAVLRGEGHWP 517
547 ARFGFAMGALPDLNODGFADVAVGAFLBEDHOGALVLYHG-TOSGVRPHPAORIAAASWP 605
518 GRFGAALTYLGDVNEKLIIDVAIGAPGEQENKAVILFHGASBSGSISSHSQRIASSQS 577
606 HALSYFGRSVDRDLDDGDLVAVAGAGAAILLSSRPVHLTSLSEVTPQAISVQORD 665
578 PRLQYFGGLSGGQDLTODGLMDLAVGAGOVULLRSLVLKGVAMRFSPEVAKAVYR 637
666 C-----RRGOEAVCLTALCFQVTSPTPGRMHQPMRTTASIDETAAGARAADGS 718
638 CMEKPSALEADATVCLTIO-----KSSLDQIGDIOSSVRFDLALDPLRLTSRAIFNET 692
719 GGRLSRRLRLSGVNTCEQLHFLHVL-DTSYLRPVALVTVALDNTTKRG-----PVYN 772
693 KNPILTRKRTLGIG-INCETLKLILPDCVEDVSVPIILHNSLVREPIPSQONLRPVLA 751
773 BGSPTSIOQLVPEFSKDCGPNCEVNDLVLOVMMDIRGSRKAPFVVGRRKVLVSTLEN 832
752 VGSQDLFTASLPEKNCQGDGLCEGD--LGVTLSFGGLT--LTVGSSLEINVLTYVNN 806
833 KRENAYNTSLSTIFSRNLHLASITFORESP-----IKVEC-AAPSAHARL-----CSYGHVP 883
807 AGEDSGYGVVSLYYPAGLSHRVSVGAOKOPHOSALRLACETVPTDEGRSSRCSVNHPI 866
884 FOTGAKVPTLFEPEFCSSLSQVFGKLTASSDSLSERNGLDENTAGTSAVIOYEPHLIF 943
867 FHGSGNGITVTFEDYSYKATLGDRLM-LMPSASSSENKASSSANTFQLEPVYKAVYTYI 925
944 SSESTLHRYEVAHPYGLTPVGPPEFKTLRLVONLGCYVVSGLIISALBPAVAGNYPLS 1003
926 SGOESTKY--FNFAISDEKMKKEAHRVRVNNLSQRDLA-ISINFWVVLINGVAVMDV 982
1004 LSOVITNNASCIYONLTERRPPVPEPELOHTNR--LNGSNTQCCVAVRCHLGOLAKGTE 1060
983 VMEASQSISPCVSEH-----KPPHSDPFLTOISRSFMLDCSIADCLQFRCVPSFSVQRE 1037
1061 VSVGLRLVHNEFFRPAKFKSLTVVSTFELGTEBEGVLOLTSARSESLLE-VYQTRRI 1119
1038 LDFTLKGNISFGWVRRTLOKRVLVVSVVAITPTDSYVSQLPGOEAFMRQOMEMVLEDEB 1097
1120 LISLWLLIGSVLGLLLALLVFLCMLKLGFF-AHKKIPEEKREE 1163
1098 VVAIPIMSSVYCALLLALLATLITATLYKGLFFKRHYKEMEDKPED 1142

```

```

STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: William Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-2
Query Match 18.5%; Score 1131.5; DB 1; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;
23 FNLDHHRRLPPGEPEAFEGYSVLOHVGCGORMLVGAPMDGPSGDRGDDYRCVGAH 82
17 FNLDVEEPTIQQ-EDAGSFGQSVOF--GGR-LVVGAPLVVAVANQGRLYDC---AAA 69
83 NAPCAKGLGYQLGNSSHPVNNHLMGMSLLETDDGPFMACAPMSAACSSVSSGIC 142
70 TGMCPRIPL-----HIREAVNMSLGLTLASTNGSRLACGPTLHRCGNSYSKSC 123
143 ARVDASFOPGSLAPTAORCP-TYWDVIVILDGNSNI--YFWEVQTFRLRLVGLFTDP 199
124 LLGSRMELIQVPRPATRECHQEMDIYFLIDGSSIDQNFQNKGVQAVMGO--PEG 181
200 EQIQVGLVOYGBSPVHNSLDFRTKEEVRBAKNLSRREGRETKYQAQIIVACTEGFSQ 259
182 TDTLFPALQVSNLKIHFPTQFRTPSPQOSLVDPVQLKGL-FTATGILTVVTLQPLNH 240
260 SHGGRPEAKLLVVTDESH-DGEELPAALKACAGVTVYGLAVLGHYLRQRQDPSF 318
241 KNGARKSAKILITVTDQKYKDPLEYSDVLPQAEKAGIRYALGV--GHAF--QGPA- 295
319 LREIRTIASDDPERFPVNTDEAALTIDVDAIGRIPLGEGSHANESSFGLMSQIGFS 378
296 ROELNTISSAPPODHVFKVDNFAALGSIQKOLEKIVAVEGTQSRASSSFQHEMSQEGFS 355
379 THRDKGILFGMGAVYDMGSSVLMLEGHRLPPRMALEDEPPALONHA-----AY 430
356 TALTDGFLGAVGSSW-----SGAFLYPPNMS-----PTFIMSOENVMRDSY 402
431 LGYSVSMWLLRGGRFLFSGAPFRHGRKVIAP-OLKKGAVRVAOSLOGEOIGSYGSGE 489
403 LGYSTELAMKGVONLVL-GAPRYOHTGKAVIFTQVSRQW--RKAAYVTGTQIGSYFGAS 459

```



```

Db 638 CMEKPSALAEADATVCLTIQ-----KSLDQDGIQSSVRDLDALDPRLTSRAIFNET 692
Qy 719 GORLSRRLRLSVGNVTCQOLHFHVL-D-TSDYLRPVALTVPALDNTTKPG-----PVLN 772
Db 693 KNPFLTRKTLGLG-IHCETLKLPLDPCVEDVSPILHLNLSVREPIPSQNLRPVLA 751
Qy 773 EGSPTSIOQLVPPSKDCGPDNCEVTDVLQVNMIDIRGSRKAPVVRGGRKVLSTTLEN 832
Db 752 VGSODLFTASLPFEKNCQGDGCEGD--LGVTLSPSGIQT--LTVGSSLEINVIIVTWN 806
Qy 833 RKENAVNTSLIIFSRNLHLASLTPORES-----IKVEC-AAPSAHARL-----CSVGHV 883
Db 807 AGEDSTIGIVASLYPAGLSHRVSGAQKQPHOSALACEVTPTEDEGRSSRCVNHPI 866
Qy 884 FOTGAKVFLFEFEFSCSSLSQVFGKLTASSDSLEENGTLOENTAOQTSAYIOYEPHLLF 943
Db 867 FHGSGNFTIVTFDVSXKATLGDRL-IMBASASSENKXSSSKATFOLELPKXAVYIMI 925
Qy 944 SEESTLHRYEHPYGTLPVGPGEPEKFTLRVONLGCYVSGLIISALLPVAHAGNYPLS 1003
Db 926 SRQESTKY--FNFATSDEKMKKEAHRVANNLSQRDLA-ISINFWVPLINGVAVMDV 982
Qy 1004 LSOVITNNAACVQNLTEPRGPVHPEELQHTNR---LNGSNTQCVVRCHIGOLAKGE 1060
Db 983 VNEAPSQSLPCVSEK-----KPPQHSDFLTQISRSPLDCLADCLQFCVDPSPSVQEE 1037
Qy 1061 VSVGLRLVHNEFPRAKFKSLTVVSTFELGTEEGSVQLTEASRWSSEILE-VYQTRPI 1119
Db 1038 LDFTLKGNISFGWAVEITLQKVLVVSVAEITFDTSYSLPQGEAFMRQMVMLEDEEV 1097
Qy 1130 LISLWILIGSVLGLLLALLVFCWKLGF-AHKKIPBEKREE 1163
Db 1098 YNAIPIIMGSSVGLLLALLATLITATLYKLGFFKRYKEMLEKPED 1142

RESULT 5
US-08-605-672-2
; Sequence 2, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Seair Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1161 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-605-672-2

Query Match 18.5%; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

Qy 23 FNLDHNRRLPPGPEAEFGYSVLOHVGCGRMVLVCAPMWGPGSGDRGDYRCVGAH 82
Db 17 FNLDVEEPTIFQ-EDAGFGGSVVOF--GGR-LVVGAPLEVVANQTRGLYDC---AAA 69
Qy 83 NAPCAKGIADYOLGNSHPAVNHLGMSLLETDDGGFMACAPLMSRACSSVPSGIC 142
Db 70 TGMQPIPL-----HIREPAVNMSLGLTLAATNGSLHLCGPIHLHVCENSISKSC 123
Qy 143 ARVDASFQPGSLAFTACRCP-TYMDVYIVLDGNSI--YPMSEVQTEFLRLVGLFIDP 199
Db 124 LLGSRWBIQTVDPATPECHQEMDIYFLIDGSGSIDQNDFNQKGFVQAVMGQ--PEG 181
Qy 200 EIOIGVLQVQESPEHMSLDFRTKEVVAAKRLSRBERETKTAQAINVACIEGFSQ 259
Db 182 TDTLPAIMOYENLKLHFTFQFRTPSPQSLVDPYOLKGL-FTATGILTVVYQLPH 240
Qy 260 SHGGRPEARLILVYVTDSEH-DEBELPAALKACEAGRVTYGIAVLGHYLRORDPSS 318
Db 241 KNGARKSKAKILIVITDQKTKDPLEISDVYPOAEKAGIIVYALGV-CHAF---OGPTA- 295
Qy 319 LREIRTTASDDPERFFNVNTDEALTIDVALGDRIFGLSGSHAENESSFGLMSQIGFS 378
Db 296 ROELNTISSAPPQOHVFVDNFFALGSIQKQLEKIVAVEGTQSRASSFQHMSQEGFS 355
Qy 379 THRLKDGILFGWAGAYDMGGSVLMLEGGHRLFPFPMALIEDEFPALQNH-----AY 430
Db 356 TALTMDDGIFLAVOSFSW-----SGGAFLYPPNMS-----PTFIMSOENVDMDRSY 402
Qy 431 LGYSVSMILNGRRRLFLSGAPRRHKGXIAF-OLKKGDAVRYAQSLOGEOISYFQSE 489
Db 403 LGYSTIELALMGVQNLV-GAPRYOHTGKAVITFOVSQW--RKAAEVTGTQIOGYFGAS 459
Qy 490 LCPIDTDRTDGTVDLVLAAPMFLGPONKETERVVYVIV--GQGSULTLQGTLOPEPPQD- 546
Db 460 LCSVDVDSGSTDILILGAPHY--EOTRGQGVVCPRLRCQRVOMOCDAVLRGEGHPW 517
Qy 547 ARFGFAMGALPDLNODGFADVAVAPLEDHOGALVYHG-TQSGVRPHPAQRIAAAMP 605
Db 518 GRFGAALTIVLQDVNEDKILDAVIGAPGEQENRGAIVYFHGASGSISSHQRIASSOLS 577
Qy 606 HALSYFGRSVYDGRDLQDGDLDVDAVAGAGAILLSSPPIHLPLPSLEVTFQAISVQRD 665
Db 578 PRLQTFGQALSGGDLTDGDLMDAVGARGVILLRSLPVTKVAGAKESVEVEYAKAYR 637
Qy 666 C-----RRRGOEAVCLTALACFOVTSRTPGRMDHOFYMFRTASLDGMTAGARAPDGS 718
Db 638 CMEKPSALAEADATVCLTIQ-----KSLDQDLDIQSSVAFDLALDGRRLTSRAIFNET 692
Qy 719 GORLSRRLRLSVGNVTCQOLHFHVL-D-TSDYLRPVALTVPALDNTTKPG-----PVLN 772
Db 693 KNPFLTRKTLGLG-IHCETLKLPLDPCVEDVSPILHLNLSVREPIPSQNLRPVLA 751
Qy 773 EGSPTSIOQLVPPSKDCGPDNCEVTDVLQVNMIDIRGSRKAPVVRGGRKVLSTTLEN 832
Db 752 VGSODLFTASLPFEKNCQGDGCEGD--LGVTLSPSGIQT--LTVGSSLEINVIIVTWN 806
Qy 833 RKENAVNTSLIIFSRNLHLASLTPORES-----IKVEC-AAPSAHARL-----CSVGHV 883

```

```

Db      807 AGEDSYGVTLVSLYPAGLSHRVSGAQKQPHOSALALACEVTEDEGRSSRCVNHPI 866
Qy      884 FQTGAQVTLLEFEFSCSSLSQVFGKLTASSDSLERNGLQNTQTSAYIOYEPHLLF 943
Db      867 FHEGSGNGFIYTFDVSYSKATLGDRL-IMRASASSENKASSSKATQLELVKYAVYVYMI 925
Qy      944 SSESILHREYVHPYGLTPVGPGPEFTTLRVQNLGCVVSGLLISALLPAVAHGNYFLS 1003
Db      926 SROESTKY--FNFAISDEKMKAEHRVNVNLSQRDLA-ISINFWVYLNGVAVMDV 982
Qy      1004 LSOVITNNASCIYQNLTEPPGPVHPEELQHTNR--LNGSNTQCQVCHLQGLAKGTE 1060
Db      983 VNEAPQSILPCVSEK-----KPPQHSDFLTQISRSPMLDCSIADCLQFRCDVPSFSQEE 1037
Qy      1061 VASGLLRLVHNEFRPAKRSKSLTVVSTFELGTESGLQLTASRSSESLLE-VYQTRPI 1119
Db      1038 LFFTLKGNLSFGVWRRTLOKVLVSVABITFTDTSVYSQLPQGEAFMRQMEVLEDEV 1097
Qy      1120 LISLWILGSLVGLGLLLALVLVCLWKLGF--AHKKIPEEKREE 1163
Db      1098 YNAIPIIMSSVGLALLLALITLTYKLGFKKRYKEMLEDKPED 1142

```

RESULT 6

US-08-482-293A-2
Sequence 2, Application US/08482293A
Patent No. 5831029

GENERAL INFORMATION:

APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bortun
STREET: 233 South Wacker Drive, 6300 Seair Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,293A
FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 37866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-293A-2

Query Match 18.5% Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4,8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

```

Qy      23 FNLDEHRLPPEPBEAFEGYSLQHVGGORWMLVGA PMWDPSGDRGRDYRCVGAH 82
Db      17 FNLDEEPTLIFQ-EDAGFGQSVQF--GGSR-LVVGAPLEVAANQGRGLDYDC--AAA 69
Qy      83 NARCAKGLDGYQLGNSSHPANNHLGMSLLETDDSGFMACAPLMSACSSVSQSGIC 142
Db      70 TCMCPILP-----HIREAVNMSLGLTASTNGSRLLACGPTLHVCGENSYSKSC 123
Qy      143 ARVDSFPOGSLAFTAPCRP-TYWDVIVLIDGSNSI--YPMSEVQTLRLRLVGLFTDP 199
Db      124 LILSRMELIQVPAPATECHQENDIVFLIDSSSIQNDPNQMKGVQVAMQO--FEG 181
Qy      200 EGIQVGLVQGESPVHENSLEDFRTKEEVNRAKNLSREGETKTAQAINVACTEFSQ 259
Db      182 TDTLPALQYSNLKIHFTEQFRTPSQSLVDPIVLKGL-TFTATGILTVVTLQPHH 240
Qy      260 SHGGRPEARLLVVTDESH--DGEELPALKACAGVTVTGIAVLGHYLRORDPSSF 318
Db      241 KNGARKSAKILLITVDQKYKDPLEYSDVLPQAKAGIRYALGV-CHAF--QGPYA- 295
Qy      319 LREIRTIASDPDERFFNVNTEAALTDIVDALGRIFGLGSHAHNESSFGLMSQIGFS 378
Db      236 ROELNTISSAPPOQHVFVNDNFALGSIQKOLEKIVAVEGTSQASSSFQHEMSQEGFS 355
Qy      379 THRLKGLIFGVNAGAYDMGSLVLEGGHRLPPRMALEDEPPALQNH-----AY 430
Db      356 TALTMDGFLGAGVFSW-----SCGAFIYPPMS-----PFIMNSQENVDMSY 402
Qy      431 LGYSVSNLILGGRRLFLISGAPRFHRRGVIAF-OLKKGAVRVAQSLQSGOISFGSE 489
Db      403 LGYSTELALMGVQNLV-LGAPRYOHTKAVIIFQVSRQW--RKKAETGTQIGSYFAS 459
Qy      490 LCPILDTDRDGTDLVLAAPMFLGPONKETERVVYVLV--GQSSILTLVQGLQPEPPD 546
Db      460 LCSVDVDDSGSTDLLIGAPHY--EOTRGQVSVCP-PRGQRYVMQCDVLRGQGHWP 517
Qy      547 ARFGFAMGALPDNLQDGFADVAAGVPLEDHQAGALYLVHG--TQSGVRHPAPRIAAAMP 605
Db      518 GRFGAALTVLGVDNEDKLIDVAGIGAQGEQRGAVYLVHGHASSESGISPSHQRISOLS 577
Qy      606 HALSYFGRSVQGRDLDDDDLVDVAVAGQAALLSSPIVHLTSLTEVTQALSVQORD 665
Db      578 PRLQYFGALSGGQDLTDGMLMDLAVGARQVLLRLSLPLKVGAVMRFSPVEVAKAYR 637
Qy      666 C-----RRRQEAVALTAALCFQVTSRTGRMDHQFYMRFSTASLDEMTAGARAAPDGS 718
Db      638 CMEERKSALEAGDAIVCLTIQ-----KSLDQGLDQGSVAFDLDLGRILSRALFNET 692
Qy      719 GQRLSPRLRLSVGNVTCEQLHFHVL-DTSYLRPVALTVPALDNTKPG-----PVLN 772
Db      693 KNPTLTRRKTGLG-ICETELKLLPDCVEVVSPIIHLNFSLVREPIPSQNLRPVLA 751
Qy      773 EGSFISIQKLVDPFSKDCGPDNECVTDVAVLVNMMDRGRSKAPFVVRGGRKRVLTLEN 832
Db      752 VGSQDLFTASLPFEKNCGQDGLCEGD--LGVTLSFGSLQOT--LTVGSLLELVNIVYVWN 806
Qy      833 KENAYNTSLGISFERNHLASLTPQRESF-----IKVEC-AAPSAHARL-----CSVGHV 883
Db      807 AGEDSYGVSLIYIYAGLSHRVSAQKQPHOSALRALACEVTEDEGRSSRCVNHPI 866
Qy      884 FQTGAQVTLLEFEFSCSSLSQVFGKLTASSDSLERNGLQNTQTSAYIOYEPHLLF 943
Db      867 FHEGSGNGFIYTFDVSYSKATLGDRL-IMRASASSENKASSSKATQLELVKYAVYVYMI 925
Qy      944 SSESILHREYVHPYGLTPVGPGPEFTTLRVQNLGCVVSGLLISALLPAVAHGNYFLS 1003
Db      926 SROESTKY--FNFAISDEKMKAEHRVNVNLSQRDLA-ISINFWVYLNGVAVMDV 982

```

QY 1004 LSOVITNNASCI VQNTLTPPPVHPEELQHTNR--LNGSNTQCVVRCHLGOLAKGTE 1060
DB 983 VMEASQSLPCVSE-----KPOHSDFLTOISRSPMLDCSIADCLQFRCVPSFVQBE 1037
QY 1061 VSVGLRLVHNEFRRAKFKSLTVSTFLGTEEGSVLQTEASRSESLLE-VVQTRPI 1119
DB 1038 LDFTLKGNLSFGWRETLOKVLVSVVAITFDTSVYSQLPGQEAEMRAQMEMVLEDEB 1097
QY 1120 LISLWILIGSVLGGLLLLALVFCIMKLGFF-AHKKIPREEKRE 1163
DB 1098 YNAIPIIMGSSVGAALLLLALITATLYKLGFFKRYKEMLEDKPED 1142

RESULT 7
US-08-943-363-2
Sequence 2, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Seair Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-2

Query Match 18.5%; Score 1131.5; DB 2; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHRLPFGPPRAERFYSVLQVGGORMLVGAPDPSGRGSDVYRCVPGAH 82
DB 17 FNLDEHRLPFGPPRAERFYSVLQVGGORMLVGAPDPSGRGSDVYRCVPGAH 82
QY 83 NACPAGKHLGVDYQLGSSHPAVNMHLGMSLLETTDGGGFMACAPLWSRAQSSVFPSSGIC 142

DB 70 TGMCPRIPL-----HIREAVNNSLGLTLASTNGSRLACGPTLHARVCGENSYSKSC 123
QY 143 ARVDASFOPOGSLAPTAQRCPTWMDVYIVLDGSNSI--YPMSEVQFLRLVGLFLIDP 199
DB 124 LLISGRMEIQTVPATPTECHQEMDIYFLIDGSSIIQNDPNQKGVQVAMQO--FEG 181
QY 200 EOIQVGLVQYGESPVHMSLSDPFRKEEVRAKNLSREGEETKTAQAINVACTEGFSQ 259
DB 182 TDTLALMOWSMLKIHFTFTQFRITSPOQSLVDPIYQLKGL-TTATGILTLVTVQLPH 240
QY 260 SHGGRPEARLLVYVTDSEH--DGEELPALKACAGAVTRYGLAVLGHYLRQRQDPSF 318
DB 241 KNGARKSAKILIVITDQKYKDPLESDVIPAOKAGIIRYALGV--EHA--QGPTA- 295
QY 319 LREIRTIASDPDERFFNVTDDEALTDIVDALGRIFLGEGSHAENESSFGLMSQIGPS 378
DB 236 ROELNTISSAPPQDHFVFNDFALGSIQKOLQEKITAVEGTQSRASSFQHEMQBGS 355
QY 379 THRLEDGILFGWAGAYDWGSLWLEGGHRLFPPRMALDEFPALQNH-----AY 430
DB 356 TALTWIDGLFLGAVGSFSW-----SGCAFILYPPNMS-----PFIMNSQENVDMDSY 402
QY 431 LGYSVSMNLGGRRLFLISGAPFRHGRKVIAF-OLKQDGAVRAQSLQGEQISYFSGE 489
DB 403 LGYSTELALMWGVQMLV--GAPRYOHTGKAVIFTOVSROW--RKAAYTGTOIGSYFGAS 459
QY 490 LCPILTDGDTDDVLVAAPMFLGPQNKETGRVYVYLV--GQOSLLTLOGTLQEPDPD- 546
DB 460 LGSVDVDSGSDTDLIGAPHY--EQTRGGQVSCPLPRQGVOMQCDAYLRGQGHFW 517
QY 547 ARFGFAMGALPDNLNODGFADVAAGAPLEDGQALYLYHG--TQSGVRPHAPQRIIAAMP 605
DB 518 GRFGALVTLGVNDEKLDIVAIAGRGQENRGAVYLFHAGSESSISFHSGRISASQLS 577
QY 606 HALSTYFGRSVNDRLLDGDIDLVDVAAGQAAILLSPPYVHLTPSLVTPQALISVQPD 665
DB 578 PRLQYFGALSGGODLTODGLMDLAVGAGVLLRLSLPLVKVAGMRFSPVEVAKAYR 637
QY 666 C-----RRRQGEAVCLTALCFQVTSRTGRMHOHYMRFSLDGMTGARAPFGS 718
DB 638 CWEERPSALEGADATVCLTIO-----KSLDQLGDIQSVFRLDLDGRILTSRAIFET 692
QY 719 GQRLSPRLRLSVGNVTCQQLHFVLD--TSDYLRVALTVFALDNTTKPG-----PVLN 772
DB 693 KNPILTRKRTIGLG--IHETIKLILPDCVEDVPSIILHNLFSLVREPIPSQNRPLVA 751
QY 773 EGSPTSIGKLVPSKDCGPDNECVTDVLQVNMDIRSKAPFVVRGGRKRVLSTLEN 832
DB 752 VGSODLFTASLPFEKNCCQDGLCEGD--LGVTLSFSGHQT--LTVGSSLEINVIYTVWN 806
QY 833 RKENAVNYSLSIIFERNHLASLTPQRESF-----IKVBC-AAPSAHNL-----CSVGHV 883
DB 807 AGEDSYGTIVSLYYPAGSHRRVSGAQKOPQOSARLACETVPFEDEGLRSRCSVNPI 866
QY 884 FQTGAKVTLFLEFEFSCSLISOVFGKLTASDSILERNGTQENTAQTSAYIOYEPHLLF 943
DB 867 FHESNGNFTIYTFDVSXATIGDM--LMRAASSENNAASSSKATFOLELPVKYAVYMI 925
QY 944 SSESTLHRYEVHPGTLFVGPGEFFKTLRVQNLGCVVVGILISALLPAVAGGNVFLS 1003
DB 926 SRQESTKY--FNFAISDEKMKKEAHRVYVNNLSQRLA--ISINFWPVLINGVAVDV 982
QY 1004 LSOVITNNASCI VQNTLTPPPVHPEELQHTNR--LNGSNTQCVVRCHLGOLAKGTE 1060
DB 983 VMEASQSLPCVSE-----KPOHSDFLTOISRSPMLDCSIADCLQFRCVPSFVQBE 1037
QY 1061 VSVGLRLVHNEFRRAKFKSLTVSTFLGTEEGSVLQTEASRSESLLE-VVQTRPI 1119
DB 1038 LDFTLKGNLSFGWRETLOKVLVSVVAITFDTSVYSQLPGQEAEMRAQMEMVLEDEB 1097
QY 1120 LISLWILIGSVLGGLLLLALVFCIMKLGFF-AHKKIPREEKRE 1163
DB 1098 YNAIPIIMGSSVGAALLLLALITATLYKLGFFKRYKEMLEDKPED 1142

```
RESULT 8
US-09-193-043-2
; Sequence 2, Application US/09193043
; Patent No. 6251395
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6251395el Human 2
; FILE REFERENCE: 27866/35004
; CURRENT APPLICATION NUMBER: US/09/193, 043
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173, 497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286, 889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362, 652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943, 363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-193-043-2

Query Match      18.5%; Score 1131.5; DB 4; Length 1161;
Best Local Similarity 29.4%; Pred. No. 4.8e-100;
Matches 348; Conservative 215; Mismatches 519; Indels 103; Gaps 40;

QY 23 FNLDEHPRLPFGPEAEFGYSVLQVHGQGMVLGAPMDPSGRGVVRCVPVGAH 82
DB 17 FNLDEEPTIFQ-EDAGGFGQSVQF--GGR-LVVGAPLEVVAHQGTLYDC---AAA 69
QY 83 NAPCAGHLGDYOLGNSHPAVNMHLMGSLLETGDCGFMACAPLMSRAGSSVFSGGIC 142
DB 70 TGMCGIPL-----HIRPAVVMISGLTLAASTNSRLLAGCPTLHRGGENSVKSGC 123
QY 143 ARVDASFQPGQSLAPTAQRCF-TYMDVIVVLGDSNSI--YPMSEVOTFLRLVYKLFIDP 199
DB 124 LLLSGWEIITQVDPDATPECPHOEMDIVFLIDSGSIDQNDPFQMGFPQAWVGQ--FFG 181
QY 200 EGIQVGLVOYGESPVHEWSLGDRTKEVYRAKUNLSRREGRTKAQIMVACTEGFQ 259
DB 182 TDTLFLMOTSNLKTHTFTQFRISPSQOSLVDPVQLKGL-TFTATGLIVVTLFHH 240
QY 260 SHGGRPEARLLVVVTDGESH--DGEELPALKACEAGVTRYGIAVLGHYLRORDPSSP 318
DB 241 KNGARSAKKILVITDQKYKDPLEYSVDIIPAEXAGIIRVAGV--GHAF--QGPTA- 295
QY 319 LREIRITASDPDERFFFNVTDEALTDIYDALCDRIFGLEGSIAENESSFGLMSQIGFS 378
DB 296 ROELNTISSAPPODHFKYDNFPAALGSIQKQOEXKYAVEGQTSRASSSFQHEMSQEGFS 355
QY 379 THRLLKGIIFGMAVGDWGSVLMWLGSHRLPEPPRALEDEPPALQNH-----AY 430
DB 356 TALTMDELGLGAVGSSW-----SGAFLYPPNMS-----FTINNQOENVDMRDY 402
QY 431 LGYSVSMILRGGRRLFLSGAPFRFRHGRKVIAP-QLKKGAVVAOSLOGEOISYFGSE 489
DB 403 LGYSTELALMKGVQNLVL--GAPRYQHTGKAVIFTOYSRQW--RKKALEVITGTOISYFGS 459
QY 490 LCPDLTDRTGTTDVLLVAAPMFLGPONKETGRVYVTV--GQOSLLTLTGTTLOPEPPD- 546
DB 460 LCSVDVDSGSDTLITIGABHY--EOTRGQGVSVCLPRGGRVQOCDAVLRGEGHPW- 517
QY 547 ARPGFMGALPDJNOGFPADVANGAPLEBHGQALVYHG--TOSGRPHPAORIAAAMP 605
DB 518 GRFGAALTVLGVDNEDKLIDVAIGAPGEBQBNRGAVYLFHGASBSGISPSHSQRIASSQS 577
```

```
QY 606 HALSYFGRSVDRDLDDDDLVDAVGAOGAAILLSSRPVHLTPSLVTPQALISVQORD 665
DB 578 PRLQYFGQLSGGQDGLTDGLMDLAVGARQGLLRSLPVLKVGVAFFSPVEVAKAYR 637
QY 666 C-----RRQGEAVCITLALCFQVTSRTPERMHGYMFTASLDWTAGARAAFDGS 718
DB 638 CWEKPSALEADATVCLTIQ-----KSLDQLOGIQSSVREDLALDQRLTSRAIFMET 692
QY 719 GQRLSPRLRLSVGNVTCQOLHFVLD--TSDFLRVALTVTFALDNTTKPG-----PVLN 772
DB 693 KNPITLTKRTLDGLG-THETLKLPLDCEVEDVSVIILHNLFSVLRPEIPSQNLRPULA 751
QY 773 EGSPTSIOKLVPFSKDCGPDNECVTDVLQVNMDIRSGRKAPFVVRGGRKVLVSTLEN 832
DB 752 VQSQDLFTASLPFEKNGCCODGLCEGD--LGVTLSFGHQF---LTVGSSLEINVIVTVWN 806
QY 833 KXENAYNLSLIFERNHLASLTQRSP-----IKVGC-AAPSAHAL-----CSVGHV 883
DB 807 AGEDSYGVVSLYYPAGSHRRVSGAQOPHQSALRLACETVPEDEGLRSRSCVNHPI 866
QY 884 FQTGAKVFLLEFEFGSSLLSQVFGKLTASDSLERNGTLOENTAOISAVIQYEPHLLF 943
DB 867 FHGSGNGFTIVTFDVSYKATLGDNM-LMRASSENNAASSKATFQLELPKXAVYTM 925
QY 944 SSESTLHRYEHPYGLTPVGPPEPKTLIRVQNLGCVYVSGLLISALLPAVAGNYFLS 1003
DB 926 SRQESTYK--FNFATSDKKMKKEAHRVYVNNLSQRDLA-ISINFVPLVINGVAVMDV 982
QY 1004 LSQVITNNASCTVQNLTPRGPVHPPELOHTNR---LNGSNTQCVVRCHGLAKTE 1060
DB 983 VNEAPSQSLPCVSEK-----KPPQHSDFLTQISRPMLDCSIADLOQRCDVPSFVQEE 1037
QY 1061 VSVGLRLRVHNEFFPRAKFSLTIVSTFELGTEGSSVQLTEASWSSESLLE-VYQTRPI 1119
DB 1038 LDTFLKGNLSRQWAFETLOKVLVSVVAEITFEDISVYQLQGEFMFAQMEMVLEBDEV 1097
QY 1120 LISLWILIGSVLGLLLALLVFCIMKLGFF-AHKRIPEEKREE 1163
DB 1098 VYAIRIIMGSSVGAALLLALITATLYKLGFPKRYKEMLEBDEP 1142

RESULT 9
US-09-688-307A-2
; Sequence 2, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688, 307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193, 043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605, 672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173, 497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286, 889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362, 652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943, 363
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-688-307A-2

Query Match      18.5%; Score 1131.5; DB 4; Length 1161;
```



```

QY 143 ARVDASFOPOGSLAPTAORCP--TYMDVIVILDGNSI--YPMSEVOTFLRLVGLFIDP 199
DB 125 LILGSRMEIITQVPDAPTECHQEMDIIVFLIDGSGSIDQNDPNQKGFVQAVMGO--FEG 182
QY 200 EOIQVGLVOYGESPVHWSLGDRTKEEVRAAKNLSRREGRTKTAQAIIVACTEGFSQ 259
DB 183 TDITFLMOYSNLKHIFFTQRTSPSQSLVDPVQLKGL--FTATGILTVVTLFPH 241
QY 260 SHGGRPEARALLVVTVDGESH--DGEELPALAKACAGAVTGYIAVLGHYLRQRDPSS 318
DB 242 KNGARKSAKILIVITDQKYKPLLESDVIPAOKAGIIRYAIGV--GHAFT--OGPTA 296
QY 319 LREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESHAENESSFGLEMSQIGFS 378
DB 297 RQELNTISSAPQDHFVKVNFALGSIQKOLEKIYAVGCTQSRASSSFQHEMSQEGFS 356
QY 379 THRUKGILFGWGAIVDMGGSVLMLEGHRLFPFRALBEFPALQNH-----AY 430
DB 357 TALTMGDLFLGAVGSSFSM-----SGGAFLYPPNMS-----PTEINMSQENVMDRDSY 403
QY 431 LGYSVSSMLRGGRFLFLSGAPFRHRGKVIAT--OLKXGAVYVAQSLQEOIGSYFSGE 489
DB 404 LGISTLALMKYQNLVL--GAPRYQHTGKAVITQVSROW--RKKAELVTGTOIGSIFGAS 460
QY 490 LCPDLTDROGTTDVLVAAPMFLGPONKETGRVYVY-----LVGQSLTLTQGLQPEPP 544
DB 461 LGSVDVDSQSTDLILIGAPHYV--EQTGQGVSVCPRLPRGRVQWCDLVRGE--QGHF 516
QY 545 QDARFPGMAGLPDLNODGFADYAVGAPLEDHGQALYLYHG--TQSGVPRHPQRILAAAS 603
DB 517 -WGRFPAALTVLGDVNEKDLIDVAIGAPGEQNRGAIVYLFHGSSEGISPSHSORIASQO 575
QY 604 MPHALSYPFGRSVGRJLDLDDDDLVDAVAGQAAILLSRPHVILPSPLEVPQALSVQ 663
DB 576 LSRRLQYFQALSGSGDILTQDGLMDLAVAGRGVLLRSLPVUKGVAKRFSVEVAKAV 635
QY 664 RDC-----RRRQEAVALTAALCFQVTSRTGRMDHOFYMRFTASLDEMTAGARAAPD 716
DB 636 YRCWEKPSALEGADATVCLTIQ-----KSLDLQDLDIQSVFDFLALDPLGRILTSRAIFN 690
QY 717 GSGQRISPRRLSVGNVTCQOLHFIYLD--TSDYLAPVALTYFALDNTTKRG-----PV 770
DB 691 ETNKPFLTRKTLGLD--HGETLKLPLPCVEDVAVPILHLNFSLVREPIBPQNLRPV 749
QY 771 LNEGSPTSIQKLVFPFGSDGCPDNECVTDVLAQVNMIDIRSGRKAFFVVRGRRKVLVSTTL 830
DB 750 LAVGSGDILTASLPFRKNCQODLCEGD--LGVTLSPSGLQT---LTVGSSLELVNIYIV 804
QY 831 ENRKENAVNTSLIIFSRLHLASLTPQRESP---IKVECAAPSAAHRL-----CSVGH 881
DB 805 WNAGEDSYGTVVSLVYPAGLSHRVSGAQKQPHQASALRLACETVPTDEGLRSSCSVNH 864
QY 882 PVPQGTAKTFLLEFEFSCSLSLQVFGKLTASDSLENGTQIQTNTAQTSAVIOEPL 941
DB 865 PLFHESSNGTFTVTFVSYKATLGDRM--LMRASASSENKASSSKATFDELEPVKXAVYT 923
QY 942 LFSSSETLRHYEVHPGTLFVPGPPEFKTLRQNLGCVVSGLLISALLPAVAGHGNF 1001
DB 924 MTSRQESSTKY--FNATSDKMKRAEHRVNNLSQSDLA--ISINFWPVULINVAW 980
QY 1002 LSLQVITNNAACIVONLTERPGRPVHPELOHTNR--LNGSNTQCVVRCGLQGLAKG 1058
DB 981 DVWMEAPSOQLPCVSR-----KPPHSDPLTQISSPMIDCSIADECLFCRCVPSFSVQ 1035
QY 1059 TEYSVULLLVNHEFRRAKFKSLTVVSTFELGTBEGSVLQLTFAARMESLLE--VQNR 1117
DB 1036 BEIDFTLKNLSFGWRETLQKKVLVSAEITFDTSVYSQLEGOAFRAQOMEWLEBD 1095
QY 1118 PILISLWILIGSVLGLLLALLLVFCMLKLGFF--AHKXIPEEKREE 1163
DB 1096 EYVNAIPITMGSSVGLLLALITATLYLKLGFFKRYKEMLEBDKPED 1142

```

```

RESULT 11
US-08-605-672-99
Sequence 99, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO.: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-99
Query Match 18.4%; Score 1126; DB 2; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1,7e-99;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;
23 FNLDEHRLPPGPEAFGYSLQHVGGGQRMVLGAPWDGSPGDRGRDYRCFVGAN 82
18 FVLDEEPTIQ--EDAGFGQSVQF--GGR--LVVGAFLVAVANQGRLYDC---AAA 70
83 NARCAKGLGQYOLGNSHRAVNNHLSLLETQDGGFMACAPLMSRACSSVSFSGIC 142
71 TGMCPIDP-----HIREAVNMSLGLTLASTNGSRLACGFLHVCGENSYSKSC 124
143 ARVDASFOPOGSLAPTAORCP--TYMDVIVILDGNSI--YPMSEVOTFLRLVGLFIDP 199
125 LILGSRMEIITQVPDAPTECHQEMDIIVFLIDGSGSIDQNDPNQKGFVQAVMGO--FEG 182
200 EOIQVGLVOYGESPVHWSLGDRTKEEVRAAKNLSRREGRTKTAQAIIVACTEGFSQ 259
183 TDITFLMOYSNLKHIFFTQRTSPSQSLVDPVQLKGL--FTATGILTVVTLFPH 241
260 SHGGRPEARALLVVTVDGESH--DGEELPALAKACAGAVTGYIAVLGHYLRQRDPSS 318
242 KNGARKSAKILIVITDQKYKPLLESDVIPAOKAGIIRYAIGV--GHAFT--OGPTA 296

```



```

Db 404 LGYSTELALMKGVQNLV-L-GAPRYOHTGKAVIFTOVSROW--RKKAENVTOIGSYFGAS 460
Qy 490 LCPDLDTRDGTDTVLVLAAPMFLGPNKENGRYVYV-----LVGOOSLTLTGTLQPPPP 544
Db 461 LCSVDVDSGSTDLILIGAPHY--EOTRGQVSVCPPLPRGRVQWCCDAVLKGE--QGHF- 516
Qy 545 ODARFGFAGALPDLNODGFADVAAGAPLEDHOGALVLYHG--TOSGVRPHPAORITAAAS 603
Db 517 -WGRFGAALTVLGDVNEKULIDVAIGAPGEBENKAVILFHGASEGSIPISSHQRITASSQ 575
Qy 604 MPHALSYFGRSVDGRDLDDGDLVDVAAGAAIILLSRPVHLTPSLVETPOAISVYQ 663
Db 576 LSPRLQYFCQALSGGODLTQDGLMDLAVAGRGVLLRSIPVLKAVNARFSEVAKAV 635
Qy 664 ROC-----RRRGQAVCLTAALCFQVTSRTGMRDHOVYMFETASLEMTAGARAARD 716
Db 636 YRWCKEPALEAGDATTCLTIO----KSLDLQGLDIOSSVAFDLALPGRITSLTAIN 690
Qy 717 GSGORLSPRLRLSVGNVTCEOLHFFVLD-TSDYLRPVALTVTYFALDNTTKG-----EV 770
Db 691 ETKNPTLTKRKLTLGLG-INCETLKLPLPCVEDVVSPIILHLNFSLVREPIPSQNLARV 749
Qy 771 LMBGPTSIQKLVPEKDCGPDNECVTDLVLYQNMDIRSKRAPVYVGRKRVLSTTL 830
Db 750 LAVGSODLFTASLPFEKNCGODLCEGD--LGVTLSFSGLQT--LTVGSSLELVNIVTV 804
Qy 831 ENRKENVATSLIIFSRNLHLASLTPORESPP-----IKVEC-AAPSAHRL-----CSVGH 881
Db 805 WNAAGEDSYGTIVSVLYPAGLSHRVSGAOKOPHQSLRLACEVTPFEDGCLSSRCSVNH 864
Qy 882 PVFOGTAKVTFLEPFESGSSLSLOVFGKLTASSDSLEBNGTLOENTAOISAVIOYEPHL 941
Db 865 PIFHEGNGTFIVTFVDSYKATLGDHM-LMRASASENNKASSKATFOLELPVKYAVYT 923
Qy 942 LBSSESTLRYEHVHPGTLPVGEGPEFKTLRQNGICVYVSSLIISALLPAVAGKXNF 1001
Db 924 MTSROESTKY--FNATSDDEKMKKEAHRVYNNLSQBDLA-ISINFVWPVLNGVAVW 980
Qy 1002 LLSQVITNNAACIVGNLTPEPPGPPVHPELQHTNR--LNGSNTQCVVRCMLGLQLAG 1058
Db 981 DVVMEAPSSQLPCVSR-----KPPQHSDELTOISHPMLDCSIADCLCPFCVPSFSQ 1035
Qy 1059 TEVSVALRLVHNEFFRAKFKSLTVVSTFELCTEGSVLQLTLEASRWSESLLE--VVQTR 1117
Db 1036 BELDFTLKNLSGWMVRETLOKKVLVVAEITFDTSVVSQLPQGAFAFRAQEMVLEED 1095
Qy 1118 PLIISIMILIGSVLGLLILALLVFCMLKLGFF-AHKTIPEEKREE 1163
Db 1096 EYVNAIPITMGSSVGLLILALLITATLTLYKLGFEKRYKEMLEBDEKPD 1142

```

RESULT 13

```

US-08-943-363-99
Sequence 99, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6500
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 99:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-99

Query Match 18.4%; Score 1126; DB 2; Length 1161;
Best Local Similarity 29.5%; Pred. No. 1,7e-99;
Matches 350; Conservative 213; Mismatches 516; Indels 108; Gaps 41;

Qy 23 FNLDEHNRLEPPGPEAFEGYSVLQHVGGGQRMVLVGAQMPDGPGRGDVYRCVGAH 82
Db 18 FNLDEHNRLEPPGPEAFEGYSVLQHVGGGQRMVLVGAQMPDGPGRGDVYRCVGAH 82
Qy 83 NAPCAKGLIDGYOLGNSHPAVNNHLGNSLLETGDDGFMACAPLISFACSSVFSGIC 142
Db 71 TGMCPRLP-----HIREAVNNISGLTLAASITNGSRLLACGPTLHRYCGENSYSKSC 124
Qy 143 ARVDASFQPGSLAPTAORCP-TYMDVITYVLDGNSI--YMWSEVQTLRLVGLKFLDP 199
Db 125 LLLSGRWELIOTVPDAPTECPHOEMDIYFLLDGSGSIDQNFQWKGVOQWMOQ--DEG 182
Qy 200 EIQVGLVQYGESPPVHEMSLDFTFKEEVBRAKXLSREGEFTYTAQAIWVACTEGFSQ 259
Db 183 TDTLFALMQYNSLNKIHFTTFQFRISPSQSLVDPVYKCL-FTTALGILTVTQLFNH 241
Qy 260 SHGGRPEARLLVVTGDESH-DGEELPALKACEAGVTRYGIAVLGHYLRQRPSSF 318
Db 242 KNGARKSAKKILIVITDQKYKDPLEYSDVIPQAEKAGIIRVAIGV--GHAF--QGPYA- 296
Qy 319 LREITIASDPERFFNVITDEBALTDIVDLGRIPLFEGSHENESSFGLMSQIGFS 378
Db 297 ROELNTISSAPPQDHVFVDNFMALGSIQKLOEKIVAVEQTQSASSSFHEMSQEGFS 356
Qy 379 THRLEDGILFGWGVYDWGSGSVLMLEGGHRLPEPPMALEDEFPPLQWHA-----AY 430
Db 357 TALTYDGLFLGAVSGFSW-----SGGAFLYPPMS-----PFIIMSGQENVDMRSY 403
Qy 431 LGYSVSMILRGRRLLFSGAPRRFRHNGKVIAF-QLKKDGAVRVAQSLQGEQIGSYFGE 489
Db 404 LGYSTELALMKGVQNLV-L-GAPRYOHTGKAVIFTOVSROW--RKKAENVTOIGSYFGAS 460
Qy 490 LCPDLDTRDGTDTVLVLAAPMFLGPNKENGRYVYV-----LVGOOSLTLTGTLQPPPP 544
Db 461 LCSVDVDSGSTDLILIGAPHY--EOTRGQVSVCPPLPRGRVQWCCDAVLKGE--QGHF- 516
Qy 545 ODARFGFAGALPDLNODGFADVAAGAPLEDHOGALVLYHG--TOSGVRPHPAORITAAAS 603
Db 517 -WGRFGAALTVLGDVNEKULIDVAIGAPGEBENKAVILFHGASEGSIPISSHQRITASSQ 575
Qy 604 MPHALSYFGRSVDGRDLDDGDLVDVAAGAAIILLSRPVHLTPSLVETPOAISVYQ 663

```



```

Db 576 LSPRLQYTGALSGGDGLTODGMDLAVGARQVILLRSLPYLKGVANRPFVBAKAV 635
Qy 664 RDC-----RRGCEAVLTALCEQVTSRRPGRMDHOFYRFFRASLDEMTAGARAAD 716
Db 636 YRCWEKPSALEAGATVCLTQ-----KSLDOLGDIOSVRFDLALDPGLTSLRAIN 690
Qy 717 GSGORLSPRRLLSYGNVTCCEOLHFFVLD-TSDYLRPVALVTYFALDNTTKG-----PV 770
Db 691 ETKNPTLIRRKTLGLG-THCELTLLKLLRPCVBDVVSPIILHNFSLVREPIPSPOKLPV 749
Qy 771 LNEGSPISIQKLVFSPSKCGDPNECVTDVLQVMDIRGSRKAPFVVRGRRKVLVSTTL 830
Db 750 LAVGSDQLFTASLPRKXNGQDGLCEGD--LGVTLSFSLQF--LTVGSSLELVITYV 804
Qy 831 ENRKEMAYNTSLISLIFSRNLHLASLTPQESF-----IKYEC-AAPBAHRL-----CSVGH 881
Db 805 WMAGEDSYGTVALSVLYPAGLSHRPVSAGAKOPHOSALRLACTCTVPFDEGLSSRCSVNH 864
Qy 882 PVFOGAKATVFLLEEFESCSLLSQVFGULTASDSLEHNGTLOENTAOYTSAYIQEPHL 941
Db 865 PIFHGSNGTFLVTDDVSFKATLGDGRM-LMRASASENNKASSKATPDLDPVKYAVYT 923
Qy 942 LFSGSTLHRVEHBYGTLVPGPGEFTKTLRVONLGCYVSGLLIISALLPVAHGNYF 100-
Db 924 MISREESTKY--FNFAISDEKMKMEABHRVYNNLSQRDLA-ISINFWPVILINGVAW 980
Qy 1002 LLSQVITNNNASCIYONLTERPGRPVPHPELGHTRN---LNSNNGQCVVRCHLQGLAKG 105-
Db 981 DVMEABPQSLPCVSE-----KPPHSDPFLQIISRPMLDOSIADCLFRCDDVBSFSQ 103-
Qy 1059 TEVSYGLRLVYNEEFRRRAKFKSLTVYSTFELGTEGSLQUTLEASRWSLELLE-VQOFR 111-
Db 1036 EELDPFLKGNLSFGWVRETLQKKVLVVSVALITPDSVYSQLPQGEAFMRQMEVLEBD 109-
Qy 1118 PILISMLIGSVGGLLALLVLCIMLGGFF-AHKRIPEEKREE 1163
Db 1096 EVNNAPIITMGSSVGLLALLATLTYLTKLGFGRHYKEMLEDKPED 1142

RESULT 14
US-09-193-043-99
Sequence 99, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395e1 Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193, 043
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286, 889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 99
LENGTH: 1161
TYPE: PRT
ORGANISM: Homo sapiens
US-09-193-043-99

```

Query Match	18.4%	Score 1126	DB 4	Length 1161
Best Local Similarity	29.5%	Pred. No. 1.7e-99		
Matches 350	Conservative 213	Mismatches 516	Indels 108	Gaps 41
QY	23	FNLDDHRLRFLPPPEALRGYSVLDHVGSGQWMLVGAPMDPSCGRDRDVRCPVGSAN	82	
		: :	: :	

Db	18	INLNDVEBPTIQ-EDAGGFGGSVVQF--GGSR-LVGAAPLEVVANQGRGLYDC---AAA	70
OY	83	NAPAKKHLDGYQLGNSSHPAVNNHKGMSLLETDDGGFMACAPLMBRACGSSVPSGTC	142
Db	71	TGMQOPIPL-----HIREAVNNMSLGLTLAASITGSRLLACGPTLHRVCGENSSKSC	124
OY	143	ARVASPOQCSLPTAQRCP-TYMDVYIVLDGNSI--YMWSEQOTLRUVGLFLIDP	199
Db	125	LLLSRBEIITVDATPECPHQEMDVLFLIDGSGSIDONDPMQKGVQAVMGO--FEG	182
OY	200	EQIOVGVLVOYESFVHEMSLDGPFRTKEBVVPAANKLSRBEGRFTKQAIIVACTEGFSQ	259
Db	183	TDLFLAMQVSNLKLHFTTFQFPITSSQGSVDPIYQKL-LFTAGILTVTQLFHH	241
OY	260	SHGRPEBARLLVVVTDEESH-DGEELPALKACEAGVTRYGIAGVILRRORPSSF	318
Db	242	KNGARKAKKILIVITDQKYKDDLEYSVDVIPAQEKAGIIRYAIGV-GHAF---QGPAT-	296
OY	319	LREIRITASDDEFFFNVTDEALTLIDVALGRITGLESBHENESSFGLMSQIGFS	378
Db	297	ROELNITISSAPQDHVKRDNFALGSIQKQLOEKIYAVETQIRASSFQHEMSQEGFS	356
OY	379	THRLKQGLIFMWAVYMGGSVLMLEGGHRLFFPRMLLEDFEPALQWHA-----AY	430
Db	357	TALTMDELFLGAVSSFSH-----SGCAFILYPRMS-----PTFIMSGENVDMRSTY	403
OY	431	LGVSSWMLRGGRRLFLSGAPRPRHRKVIAT-QLKKDGAVRVAQSLQSGISYFGSE	489
Db	404	LGYSTELALMKGVQNLVL-GAPRYQHTGKAVIFQVSRQW--RKALEVTVGQISYFGAS	460
OY	490	LCPLDTPBDGTTDLVLAAPMFLCPQNKETRRVVY-----LVQCGSLITLQGTLOPEPP	544
Db	461	LCASVDVSDGSTDLILGAPRY--EOTRRQGVAVCP,PRGRVOMOCDAVLRG-OGHP-	516
OY	545	QDAFFGFAMGLPYLNDGPFADVAVGAPLEBEGHAGALYLYVG-TOSGVRPHAPORIAAS	603
Db	517	-WGFAGALTYLGVNDEKLDIVAIAGBEGENNGAYILFFGASSEGISPSHSQRIASSQ	575
OY	604	MPHALSYFGSRVSDRLDDGDDLVDAVVGAGGAAILLSSREIVHLTPSLETPQASIVQ	663
Db	576	LSPLRQFGQALSGGQDULDGLMDLAVGARQVILLRSLPVLKVGAMRSPSVEAAV	635
OY	664	RDC-----RRRQEAVALCLTALCFQVTSFTGPRMWHQFYMRFTASIDENFTAGARAAD	716
Db	636	YRCWEERPSALEADATVCLTIQ-----KSLDQDGIQSSVRFDLALDPELTSRAIFN	690
OY	717	GSGRSLRRLRLSVGNVTCQULFHVLD-TSDVLRPALVTVTALDNTTRG-----PV	770
Db	691	ETKNPULTRRRRTIGLG-IHCETLLPLLPDCEVDVSPILLHLNLSLREPIPSQNLRPV	749
OY	771	LNESGPTSIQKLVPFSCDGPDNCEYDVLVQVMMDIRGSRKAPFVVRGGRKVLVSTL	830
Db	750	LAVSSQDLFTASLPFEKNCQDGLCEBD--LGVTLSSGLQI--LTVGSSLELNVITV	804
OY	831	ENRKENAYNTSLSIIFSRNLHLASLTPQRESF---IKVEC-AAPSAHARL-----CSVGH	881
Db	805	WNAGEDYGTVALSYYPAGLSHRVSAQOQPHQSALRIACETVPTEDEGLRSSRCVNH	864
OY	882	PVPOFGAKVTFLEFEFSCSLSQVGVKLTASDSLSBRNGTLOENTRAQTSAYIQYEH	941
Db	865	PIFHEGNGFTIVFDVSYKATLGDRLM-LMRASASSENNKAASSKATFOLELPKYAAVT	923
OY	942	LFESESTLHREYVPGTLPVGPBEPFKTTLRYONLCGVYVSGIISALPLVAVHAGNYF	1001
Db	924	MISQOESTKY--FNFAISDEKKKKEKHEKRYVNNLSORDLA-ISINFVVLINGAVW	980
OY	1002	LSLSQVITNNASCIQVNLTEBPGBPVHAPBEOHTNR--LNGSNTQOCVVRCHLGOLAKG	1058
Db	981	DVWAEAPSQSILPCYSEF-----KPGQSDFLQTSRBPMLDCSTADCLQFCRDVPSPVQ	1035
OY	1059	TEVSVGLLRVLVHNEFFPRAKFSLTVVSTELGTBESSVIQLTASWMSBLE-VVQTR	1117
Db	1036	EBDLPFTLKGNI,SPFWAVETLQKVLVYVAEITFDITSVQLPQGBAFMRQOMVMVLEED	1095

981 DYVNEAPDSQSLPCVSEK-----KPGQSDFTLTQTSRSPMLDCSTADLTQGFCDVPSFSVQ 1035
 1059 TEVSVGLRLVHNHFFPRAKFKSLTVVSTFELGTBEESVYQLTASRMSESLLE-VYQTR 1117
 1036 EELDTFLTKGNISFGWAEVLQKLYLVVVAEITFDITSVYSQLPGQAEFMKRAMQWVLEED 1095

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 10.1646 Seconds

(without alignments)
1872.646 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDASFQSGSLAPFA.....FLREIRITASDPDERFFENV 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	473	46.1	272	2 A55348	integrin alpha-1 -
2	467.5	45.6	1151	2 A45226	integrin alpha-1 c
3	465.5	45.4	1180	2 A35854	integrin alpha-1 c
4	454	44.3	1178	2 S44142	VLA-2 protein homo
5	451	44.0	1170	2 I45914	integrin alpha 2 s
6	441	43.0	1181	2 A33998	integrin alpha-2 c
7	336	32.8	191	2 I47230	VLA-2 protein - pi
8	227.5	22.2	1747	1 A45974	collagen alpha 1(X
9	227.5	22.2	1857	2 S31212	collagen alpha 1(X
10	227.5	22.2	1888	2 S78476	collagen alpha 1(X
11	227.5	22.2	3124	1 A40020	collagen alpha 1(X
12	214	20.9	493	2 A33809	cartilage matrix p
13	205	20.0	496	2 A37979	cartilage matrix p
14	204.5	20.0	929	2 I51027	type XII collagen
15	201.5	19.7	3137	1 A37997	collagen alpha 3(V
16	199	19.4	3051	2 S42373	hypothetical prote
17	194.5	19.0	500	2 S66522	cartilage matrix p
18	177.5	17.3	1179	2 A53213	integrin alpha-E c
19	176.5	17.2	3176	1 CGHJ3A	collagen alpha 3(V
20	175	17.1	724	2 A48569	antigen Em100 - Et
21	170.5	16.6	843	2 A40970	undulin 1 - human
22	169	16.5	741	2 T46488	hypothetical prote
23	158.5	15.5	1153	1 RWHU1B	cell surface glyco
24	158	15.4	1163	2 RWHU1B	lymphocyte functio
25	157.5	15.4	1153	2 S00551	leukocyte surface
26	156	15.2	550	2 T23760	hypothetical prote
27	156	15.2	2944	2 A54849	collagen alpha 1(V
28	154	15.0	2813	1 VWHU	von Willebrand fac
29	153.5	15.0	1163	1 RWHU1C	cell surface glyco

30	152	14.8	1170	2 S03308	cell surface glyco
31	142.5	13.9	712	2 A45638	immunodominant m
32	141	13.8	341	2 T32949	hypothetical prote
33	136.5	13.3	427	2 G00039	von Willebrand fac
34	129.5	12.6	414	2 P50323	von Willebrand fac
35	123	12.0	1286	2 A88396	protein M01E10.2 (
36	116.5	11.4	13055	2 T16580	hypothetical prote
37	114.5	11.2	567	2 T28797	hypothetical prote
38	112.5	11.0	661	2 T16597	hypothetical prote
39	110	10.7	643	2 T19549	hypothetical prote
40	107	10.4	449	2 T33257	hypothetical prote
41	103.5	10.1	2098	2 T18397	protein CTRP - mal
42	100	9.8	1029	1 S21369	collagen alpha 2(V
43	97.5	9.5	780	2 A34102	von Willebrand fac
44	95.5	9.3	1019	1 A32856	collagen alpha 1(V
45	95	9.3	445	2 B40970	undulin 2 - human

ALIGNMENTS

RESULT 1

A55348
integrin alpha-1 - chicken (fragment)
C/Species: Gallus gallus (chicken)
C/Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
C/Accession: A55348
R/Kern, A.; Briesewitz, R.; Bank, I.; Marcantonio, E.E.
J. Biol. Chem. 269, 22811-22816, 1994
A/Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1
A/Reference number: A55348; MUID:94357930; PMID:7521332
A/Accession: A55348
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-272 <KER>
A/Cross-References: GB:U0114
C/Suprafamily: unassigned collagens; von Willebrand factor type A repeat homology
F/55-230/Domain: von Willebrand factor type A repeat homology <WMA2>

Query Match
Best Local Similarity 46.1%; Score 473; DB 2; Length 272;
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

Qy	1	GICARVDASFQSGSLAPFAQRCPTVMVIVLDGNSIYPWSEVQTFRLRLVGLFIDP	60
Db	30	GVCSNVSTFERNKAVASVQCKTDIVIVLDGNSIYPWSEVQTFRLRLVGLFIDP	89
Qy	61	FOIQVGLVQSGSPVHNSLDGFTKEEVRAAKNLSRREGRETKTAQAINVACTEFSQ	120
Db	90	OOTQVGIQVQGVVHFEVLYNTSTTEVMDALRIQRGGTQWTALGIDTAREAEATE	149
Qy	121	SHGGRPEARLLVVTDSHDEGLPALAKCEGRTRGCIANLGHYLRQRDPSSFL	180
Db	150	AHGARGVQKVVIVTGDESHVNLQEVLDKCEDENIQRPAIALTGSYSGNLSTKFEV	209
Qy	181	REIRITASDPDERFFENV	198
Db	210	EEIKSLAKPTKEKHFNV	227

RESULT 2

A45226
integrin alpha-1 chain - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2000
C/Accession: A45226
J. Briesewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A/Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A/Reference number: A45226; MUID:93155124; PMID:8428973
A/Accession: A45226
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA

A:Residues: 1-1151

 A:Experimental source: hepatoblastoma cell line HepG2
 A>Note: sequence extracted from NCBI backbone (NCBI:P:124326)
 C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
 F:142-317/Domain: von Willebrand factor type A repeat homology <VMA1>

Query Match 45.6%; Score 467.5; DB 2; Length 1151;

Best Local Similarity 46.5%; Pred. No. 5.9e-35;

Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

1 GICARVDASFOGSLAPTAQRCPTVMGVVILDGNSIYPMSVQTFRLRLVGLFLIDP 60

118 GICSDVSPFQVYVNSIAP-VQECSTOLDIVILDGNSIYPMDSVTAFLNDLKRMDIGP 176

61 EOIQVGLVOYGESPVHEMSLGFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 120

177 KQIQVGLVOYGEVNTHEFNLNKYSSTEEVLPAAKKIVQRGQRMTALGTDTRAKKAFTE 236

121 SHGGRPEARLLVVTVDGSHDEELPAALKACEAGRTVRYGIAVLGHYLRQRDPSSFL 180

237 ARGARGVKKWVIVTDGSHDNHRLKKVIOCEDEENIQFSAIIIGSVNRGNLSTEKTV 296

181 REIRTIASDPDERFFNFV 198

297 EEIKSIASEPTEKHFNFV 314

RESULT 3

A:Residues: 1-1180 <IGN>

A:Cross-references: GB:X52140; NID:G56493; PIDN:CAA36384.1; PID:G56494

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

C:Keywords: cell adhesion; cytoskeleton; transmembrane protein

F:170-345/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 45.4%; Score 465.5; DB 2; Length 1180;

Best Local Similarity 46.5%; Pred. No. 9.3e-35;

Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

1 GICARVDASFOGSLAPTAQRCPTVMGVVILDGNSIYPMSVQTFRLRLVGLFLIDP 60

146 GICSDVSPFQVYVNSIAP-VQECSTOLDIVILDGNSIYPMDSVTAFLNDLKRMDIGP 204

61 EOIQVGLVOYGESPVHEMSLGFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 120

205 KQIQVGLVOYGEVNTHEFNLNKYSSTEEVLPAAKKIVQRGQRMTALGTDTRAKKAFTE 264

121 SHGGRPEARLLVVTVDGSHDEELPAALKACEAGRTVRYGIAVLGHYLRQRDPSSFL 180

265 ARGARGVKKWVIVTDGSHDNHRLKKVIOCEDEENIQFSAIIIGSVNRGNLSTEKTV 324

181 REIRTIASDPDERFFNFV 198

325 EEIKSIASEPTEKHFNFV 342

RESULT 4

A:Residues: 1-1170 <KAM>

A:Cross-references: GB:L25886; NID:G439695; PIDN:AAB59255.1; PID:G439696

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

C:Keywords: cell adhesion; cytoskeleton; transmembrane protein

F:161-336/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 44.4%; Score 451; DB 2; Length 1170;

Best Local Similarity 44.4%; Pred. No. 2e-33;

Matches 88; Conservative 37; Mismatches 73; Indels 0; Gaps 0;

R:Edelman, J.M.; Chan, B.W.; Uniyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze

submitted to the EMBL Data Library, January 1994

A:Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not

A:Reference number: S44142

A:Accession: S44142

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1178 <EDB>

A:Cross-references: EMBL:229987; NID:G473098; PIDN:CAA82877.1; PID:G473099

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

F:169-344/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 44.3%; Score 454; DB 2; Length 1178;

Best Local Similarity 46.0%; Pred. No. 1.1e-33;

Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;

1 GICARVDASFOGSLAPTAQRCPTVMGVVILDGNSIYPMSVQTFRLRLVGLFLIDP 60

144 GICSDVSPDFQVYVNSIAP-VQECSTOLDIVILDGNSIYPMDSVTAFLNDLKRMDIGP 203

61 EOIQVGLVOYGESPVHEMSLGFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 120

204 KQIQVGLVOYGEVNTHEFNLNKYSSTEEVLPAAKKIVQRGQRMTALGTDTRAKKAFTE 263

121 SHGGRPEARLLVVTVDGSHDEELPAALKACEAGRTVRYGIAVLGHYLRQRDPSSFL 180

264 TSGGRPGATKVVVVTVDGSHDGSGLKTVIQCNNDILRFQIAVLGYLNRAALDTKULI 323

181 REIRTIASDPDERFFNFV 198

324 KEIKAIASIPTEKHFNFV 341

RESULT 5

A:Residues: 1-1170 <KAM>

A:Cross-references: GB:L25886; NID:G439695; PIDN:AAB59255.1; PID:G439696

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology

C:Keywords: cell adhesion; cytoskeleton; transmembrane protein

F:161-336/Domain: von Willebrand factor type A repeat homology <VMA2>

Query Match 44.0%; Score 451; DB 2; Length 1170;

Best Local Similarity 44.4%; Pred. No. 2e-33;

Matches 88; Conservative 37; Mismatches 73; Indels 0; Gaps 0;

1 GICARVDASFOGSLAPTAQRCPTVMGVVILDGNSIYPMSVQTFRLRLVGLFLIDP 60

136 GICSDVSPDFQVYVNSIAP-VQECSTOLDIVILDGNSIYPMDSVTAFLNDLKRMDIGP 195

61 EOIQVGLVOYGESPVHEMSLGFRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQ 120

196 TKTQGLIQYANPFRVFNLTFSKDEMIKTSOTFOYGDLTWTFPAIQYARTAVST 255

121 SHGGRPEARLLVVTVDGSHDEELPAALKACEAGRTVRYGIAVLGHYLRQRDPSSFL 180

256 AAGGRPGATKVVVVTVDGSHDGSGLKTVIQCNNDILRFQIAVLGYLNRAALDTKULI 315

181 REIRTIASDPDERFFNFV 198

316 KEIKAIASIPTEKHFNFV 333

RESULT 6

Db 309 PDSTHYVNV 317

RESULT 11

A40020
collagen alpha 1(XII) chain precursor - chicken

N:Alternate names: fibroinmerin

C:Species: Gallus gallus (chicken)

C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Jan-2001

C:Accession: A40020, A34485, B34485, A28037, S22814, S22254, S28811

R:Yamagata, M., Yamada, K.M., Yamada, S.S., Shinomura, T., Tanaka, H., Nishida, Y., Obata, J., Cell Biol. 115, 209-221, 1991

A:Title: The complete primary structure of type XII collagen shows a chimeric molecule with a novel region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site.

A:Reference number: A40020, MUID:92011862, PMID:1518137

A:Accession: A40020

A:Molecule type: mRNA

A:Residues: 1-3124 <YAM>

A:Cross-references: GB:D00824, NID:g222810, PIDN:BA00701.1, PID:g222811

A>Note: In the authors' translation residues 1216-1219 are shown after residue 1235 and.

R:Gordon, M.K., Gerecke, D.R., Dublet, B., van der Rest, M., Olsen, B.R.

J. Biol. Chem. 264, 19772-19778, 1989

A:Title: Type XII collagen. A large multidomain molecule with partial homology to type I

A:Reference number: A34485, MUID:90062079, PMID:258192

A:Accession: A34485

A:Molecule type: mRNA

A:Residues: 2456-2758, 'A', 2760-2802, 'F', 2804-2976, 'F', 2978-3124 <GOR>

A:Cross-references: EMBL:J05137, NID:g211284, PIDN:AAA48635.1, PID:g211285

A:Accession: B34485

A:Molecule type: protein

A:Residues: 2772-2792/2846-2873 <GOR2>

R:Gordon, M.K., Gerecke, D.R., Olsen, B.R.

Proc. Natl. Acad. Sci. U.S.A. 84, 6040-6044, 1987

A:Title: Type XII collagen: distinct extracellular matrix component discovered by cDNA c

A:Reference number: A28037, MUID:87317590, PMID:3476925

A:Accession: A28037

A:Molecule type: mRNA

A:Residues: 2960-2976, 'F', 2978-3074, 'AG', <GOR3>

A:Cross-references: EMBL:M1375, NID:g211649, PIDN:AAA48718.1, PID:g211650

A>Note: this sequence has been revised in reference A34485

R:Koch, M., Bernasconi, C., Chiquet, M.

Eur. J. Biochem. 207, 847-856, 1992

A:Title: A major oligomeric fibroblast proteoglycan identified as a novel large form of

A:Reference number: S23814, MUID:92362621, PMID:1323460

A:Accession: S23814

A:Molecule type: protein

A:Residues: 'X', 1333, 'C', 1335-1347, 1914-1928, 2504, 'X', 2506, 'X', 2508-2511, 'X', 2513-2517 <

R:Dublet, B., van der Rest, M.

J. Biol. Chem. 262, 17724-17727, 1987

A:Title: Type XII collagen is expressed in embryonic chick tendons. Isolation of pepsin

A:Reference number: S22254, MUID:88087065, PMID:3121603

A:Accession: S22254

A:Molecule type: protein

A:Residues: 2831-2832, 'T', 2834, 'R', 2836-2843, 3002-3014 <DDB>

R:Trueb, J., Trueb, B.

Biochim. Biophys. Acta 1171, 97-98, 1992

A:Title: The two splice variants of collagen XII share a common 5' end.

A:Reference number: S28811, MUID:93042014, PMID:1420368

A:Accession: S28811

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24, 1189-1257, 'S', 1259-1263, 'E', 1265-1280 <TRU>

A:Cross-references: EMBL:X67327

C:Genetics:

A:Introns: 2845/3, 2863/3, 2887/3, 2899/3, 2922/1, 2985/1, 3008/1, 3065/1

C:Superfamily: collagen alpha 1(XII) chain; fibronectin type III repeat homology; von Willebrand factor type I repeat homology; cell binding; coiled coil; connective tissue; disulfide bond; signal sequence; status predicted <SIG>

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-3124/Product: collagen alpha 1(XII) chain #status predicted <MAT>

F:24-1189-3124/Product: collagen alpha 1(XII) chain short splice form #status predicted

F:24-114/Domain: IIIA #status predicted <IIIA>

F:24-105/Domain: fibronectin type III repeat homology <FN3A>

F.1137-301/Domain: von Willebrand factor type A repeat homology <vWA1>
 F.1332-425/Domain: IIIB #status predicted <IIIB>
 F.332-425/Domain: IIIB #status predicted <IIIB>
 F.332-414/Domain: fibronectin type III repeat homology <FN3B>
 F.437-601/Domain: von Willebrand factor type A repeat homology <vWA2>
 F.629-1178/Domain: IIIC #status predicted <IIIC>
 F.630-711/Domain: fibronectin type III repeat homology <FN3C>
 F.721-802/Domain: fibronectin type III repeat homology <FN3D>
 F.812-895/Domain: fibronectin type III repeat homology <FN3E>
 F.905-966/Domain: fibronectin type III repeat homology <FN3F>
 F.995-1076/Domain: fibronectin type III repeat homology <FN3G>
 F.1086-1169/Domain: fibronectin type III repeat homology <FN3H>
 F.1197-1361/Domain: von Willebrand factor type A repeat homology <vWA3>
 F.1384-2295/Domain: IIID #status predicted <IIID>
 F.1384-1465/Domain: fibronectin type III repeat homology <FN3I>
 F.1474-1557/Domain: fibronectin type III repeat homology <FN3J>
 F.1566-1647/Domain: fibronectin type III repeat homology <FN3K>
 F.1655-1738/Domain: fibronectin type III repeat homology <FN3L>
 F.1756-1838/Domain: fibronectin type III repeat homology <FN3M>
 F.1847-1928/Domain: fibronectin type III repeat homology <FN3N>
 F.1937-2019/Domain: fibronectin type III repeat homology <FN3O>
 F.2028-2110/Domain: fibronectin type III repeat homology <FN3P>
 F.2119-2199/Domain: fibronectin type III repeat homology <FN3Q>
 F.2207-2294/Domain: fibronectin type III repeat homology <FN3R>
 F.2325-2490/Domain: von Willebrand factor type A repeat homology <vWA4>
 F.2438-2440/Domain: cell adhesion #status predicted
 F.2509-2750/Domain: IXP homologous to NC4 domain of type IX collagen #status predicted
 F.2751-2902/Domain: collagenous COL2 #status predicted <COL2>
 F.2899-2901/Domain: cell attachment (R-G-D) motif
 F.2903-2945/Domain: non-collagenous NC2 #status predicted <NC2>
 F.2946-3048/Domain: collagenous COL1 #status predicted <COL1>
 F.3049-3124/Domain: non-collagenous NC1 #status predicted <NC1>
 F.332.1006.1032.1044.1512.1767.2210.2273.2532.2683/Binding site: carboxylate (Asn) (covar
 F.2780.2789.2836.2842.2860.2869.3004.3007/Modified site: Hydroxyproline (Pro) #stat

Query Match 22.2%; Score 227.5; DB 1; Length 3124;
 Best Local Similarity 34.1%; Pred. No. 3.3e-12;
 Matches 59; Conservative 32; Mismatches 67; Indels 15; Gaps 5;

Oy 28 DVAIVLDGNSI--YPMSEVQTFRLRLVGLFLDPEQIQVGLVQVGBSPVHEWSLGDFT 85
 Db 1199 DIVLVVGVGSSISRPNKFTYRNFISRIVEFVDIGDPKVQIGLQVQSGDPTFNNLNAVRT 1258
 Oy 86 KEVVRAKRLSRREGFTTAQIMWACREGSSQSHGGRPAARLLVVTDESDGEE 145
 Db 1259 KEALLDVNTLPYK--GGNTLTGMALDFILKNRKQEGARPRARKIGVLLTDDKSDDDV 1317
 Oy 146 LPALAKACGAGRTYRGIAVLGHYLRQRPDPSFLEIRITASDPERPFNV 198
 Db 1318 TPSSRLRDEG--VELVAIGI-----KNADEN---ELKQIATDPDDIHAYNV 1358

RESULT 12
 A33809
 Cartilage matrix protein precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #ext_change 21-Jul-2000
 C:Accession: A33809; A26364
 R:Kiss, I.; Deak, F.; Holliday Jr., R.G.; Delius, H.; Mebus, K.A.; Fritmberger, E.; Argir
 J. Biol. Chem. 264, 8126-8134, 1989
 A:Title: Structure of the gene for cartilage matrix protein, a modular protein of the ex
 trinsic, von Willebrand factor, complement factors B and C2, and epidermal growth factor.
 A:Reference number: A33809; MUID:89255246; PMID:2542265
 A:Accession: A33809
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-493 <KIS>
 A:Cross-references: GB:X12346; GB:X12347; GB:X12348; GB:X12349; GB:X12350; GB:X12351; GB
 R:Argaves, W.S.; Deak, F.; Sparks, R.J.; Kiss, I.; Goettink, P.F.
 P:Proc. Natl. Acad. Sci. U.S.A. 84, 464-468, 1987
 A:Title: Structural features of cartilage matrix protein deduced from cDNA.
 A:Reference number: A26364; MUID:87092429; PMID:3025875
 A:Accession: A26364
 A:Molecule type: mRNA

Query Match	19.7%	Score 201.5	DB 1	Length 3137
Best Local Similarity	29.8%	Pred. No. 8.6e-10		
Matches 59, Conservative	42	Mismatches 72	Indels 25	Gaps 6

Search completed: July 16, 2003, 07:56:47
Job time : 12.1646 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 5.1519 Seconds

(without alignments)
1594.036 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025
Sequence: 1 GICARVDASFOQSGIAPTA.....FLREIRITASDPDERFFENV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1025	100.0	1167	ITAG_HUMAN	O75578 homo sapien
2	544	53.1	1189	ITAH_HUMAN	O90KX5 homo sapien
3	473	46.1	1285	ITAI_CHICK	O90615 gallus gall
4	467.5	45.6	1151	ITAI_HUMAN	P56199 homo sapien
5	465.5	45.4	1180	ITAI_RAT	P18614 rattus norv
6	454	44.3	1178	ITAI_MOUSE	O62469 mus musculu
7	451	44.0	1170	ITAI_BOVIN	P53710 bos taurus
8	441	43.0	1181	ITAI_HUMAN	P17301 homo sapien
9	227.5	22.2	1888	CAIC_CHICK	P13944 gallus gall
10	227.5	22.2	3124	CAIC_CHICK	P32018 gallus gall
11	214	20.9	493	CAMA_CHICK	O50599 gallus gall
12	213.5	20.8	3119	CAIC_MOUSE	O60847 mus musculu
13	209.5	20.4	3063	CAIC_HUMAN	O99715 homo sapien
14	205	20.0	496	CAMA_HUMAN	P21941 homo sapien
15	204.5	20.0	929	CAIC_NOTVI	O91145 notophthalm
16	201.5	19.7	3137	CA36_CHICK	P15889 gallus gall
17	199	19.4	3767	MUA3_MOUSE	P34576 caenorhabdi
18	196	19.1	956	MTN2_MOUSE	O08746 mus musculu
19	194.5	19.0	500	CAMA_MOUSE	P51942 mus musculu
20	191	18.6	956	MTN2_HUMAN	O00339 homo sapien
21	186	18.1	624	MTN4_MOUSE	O08929 mus musculu
22	185.5	18.1	619	MTN4_HUMAN	O95460 homo sapien
23	177.5	17.3	1179	ITAE_HUMAN	P38570 homo sapien
24	176.5	17.2	3176	CA36_HUMAN	P12111 homo sapien
25	169.5	16.5	2482	VWF_PIG	O28833 sus scrofa
26	164	16.0	2813	VWF_CANFA	O28835 canis fami
27	160	15.6	1152	ITAD_HUMAN	Q13349 homo sapien
28	158.5	15.5	1152	ITAD_HUMAN	Q13349 homo sapien
29	158	15.4	1163	ITAI_MOUSE	P24063 mus musculu
30	157.5	15.4	1153	ITAI_MOUSE	P05555 mus musculu
31	156	15.2	2944	ITAE_HUMAN	O02188 homo sapien
32	154.5	15.1	1167	ITAE_MOUSE	O60677 mus musculu
33	154	15.0	2813	VWF_HUMAN	P04275 homo sapien

34	153.5	15.0	1163	ITAX_HUMAN	P20702 homo sapien
35	152	14.8	1170	ITAI_HUMAN	P20701 homo sapien
36	147	14.3	486	MTN3_HUMAN	O15232 homo sapien
37	143	14.0	452	MTN3_CHICK	O42401 gallus gall
38	140	13.7	547	COCH_CHICK	O42163 gallus gall
39	135	13.2	481	MTN3_MOUSE	O35701 mus musculu
40	120.5	11.8	550	COCH_HUMAN	O43405 homo sapien
41	120.5	11.8	552	COCH_MOUSE	O62507 mus musculu
42	110.5	10.8	639	CA1C_RABIT	O28902 oryctolagus
43	100	9.8	1029	CA26_MOUSE	O02788 mus musculu
44	95.5	9.3	1019	CA16_CHICK	P20785 gallus gall
45	92	9.0	1018	CA26_HUMAN	P12110 homo sapien

ALIGNMENTS

```

RESULT 1
ITAG_HUMAN          STANDARD;          PRT; 1167 AA.
ID   ITAG_HUMAN
AC   O75578; O9UH28;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-UN-2002 (Rel. 41, Last annotation update)
DE   Integrin alpha-10 precursor.
OS   ITGA10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Articular chondrocytes;
RX   MEDLINE=98352078; PubMed=9685391;
RA   Camper L., Hellman U., Lundgren-Akerlund E.;
RT   "Isolation, cloning, and sequence analysis of the integrin subunit
RT   alpha10, a beta1-associated collagen binding integrin expressed on
RT   chondrocytes."
RL   J. Biol. Chem. 273:20383-20389(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart, and Endothelial cells;
RX   MEDLINE=20169197; PubMed=10702680;
RA   Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
RA   Wang S.-X., Langley R., Kristensen G.W.;
RT   "The integrin alpha10 subunit: expression pattern, partial gene
RT   structure, and chromosomal localization."
RL   Cytogenet. Cell Genet. 87:238-244(1999).
CC   - FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC   - SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC   ASSOCIATES WITH BETA-1.
CC   - SUBCELLULAR LOCATION: Type I membrane protein.
CC   - TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN
CC   MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.
CC   - DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC   WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC   - SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC   - SIMILARITY: CONTRAINS 1 VWFA DOMAIN.
CC   - SIMILARITY: CONTRAINS 7 FG-GAP REPEATS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by, and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; AF074015; AAC31952.1; -
DR   EMBL; AF112345; AAF21944.1; -
DR   EMBL; AF172723; AAF61638.1; -
DR   HSSP; PI7301; IAOX.
DR   Genew; HGNC:6135; ITGA10.

```

	Query Match	100.0%;	Score 1025;	DB 1;	length 1167;	
	Best Local Similarity	100.0%;	Pred. No. 7,6e-85;			
Matches	198;	Conservative	0;	Mismatches	0;	Indels 0; Gaps 0
QY	1	GICARVNDSPFOQGSLAPTAORCPPTMYDVIVILDGNSIYPMSSEVOTFLRLVGKLFIDP	60			
Db	140	GICARVDSPFOQGSLAPTAORCPPTMYDVIVILDGNSIYPMSSEVOTFLRLVGKLFIDP	199			
QY	61	EQIQVLVOYGESPHEWISLSDPRTKEEYVRAAKNLSRREGRETAKQAIMVACTEGFSQ	120			
Db	200	EQIQVLVOYGESPVHEMSLSDPFRTEEYVRAAKNLSRREGRETAKQAIMVACTEGFSQ	259			
QY	121	SHGGPEPARLLVVTTDGDSSHGEELPALKACEAGRVTIRYGIAYLGHLRRORDPSSTL	180			
Db	260	SHGGPEPARLLVVTTDGDSSHGEELPALKACEAGRVTIRYGIAYLGHLRRORDPSSTL	319			
QY	181	REIRTIASDPDERFFENV 198				
Db	320	REIRTIASDPDERFFENV 337				
RESULT 2						
ITAH_HUMAN						

```

ID AD ITAH HUMAN STANDARD; PRT; 1189 AA.
AC Q9UKX5; O9UK01.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGA11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
XX [1]
XX SEQUENCE FROM N.A.
XX TISSUE=Fetal heart;
XX MEDLINE=99417678; Pubmed=10466209;
RA Lenneth K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D.,
RA Wang S.-X., Morris C.M., Krissansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel
RT human integrin alphai1 subunit (ITGA11).";
RL Genomics 60:179-187(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal muscle, and Uterus;
RC MEDLINE=99395147; Pubmed=10464311;
RX Velling T., Kusche-Gullberg M., Sejersten T., Gullberg D.;
RA "cDNA Cloning and Chromosomal Localization of Human Alpha(11)
RT Integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin A-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
RN [3]
RN SEQUENCE OF 954-1188 FROM N.A.
RP TISSUE=Fibroblast;
RC Andreu N., Estivill X., Becarcelier M., Smoy L.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC FETAL MUSCLE CELLS (IN VITRO).
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF109681; AAF01258.1; -
DR EMBL; AF137378; AAD51919.2; -
DR EMBL; AL359064; CAB94392.1; -
DR HSSP; P17301; 1A0X.
DR Genew; HGNC:6136; ITGA11.
DR MIM; 604789; -
DR InterPro; IPRO00413; Integrin_alpha.
DR InterPro; IPRO02035; VFMA_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRIN.
DR PRINTS; PR00453; VFMADOMAIN.
DR SMART; SM00191; Int_alpha; 5.

```

```

DR SMART, SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS00234; VWFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 1 1189
FT DOMAIN 23 1189
FT TRANSMEM 1143 1165
FT DOMAIN 1166 1189
FT REPEAT 38 94
FT REPEAT 102 163
FT REPEAT 167 345
FT REPEAT 359 420
FT REPEAT 422 475
FT REPEAT 477 537
FT REPEAT 539 598
FT REPEAT 601 653
FT DOMAIN 1154 1162
FT DOMAIN 1174 1177
FT CA_BIND 488 496
FT CA_BIND 551 559
FT CA_BIND 613 621
FT DISULFID 76 83
FT DISULFID 121 139
FT DISULFID 129 159
FT DISULFID 659 668
FT DISULFID 674 729
FT DISULFID 781 787
FT DISULFID 881 893
FT CARBOHYD 82 82
FT CARBOHYD 95 95
FT CARBOHYD 291 291
FT CARBOHYD 331 331
FT CARBOHYD 338 338
FT CARBOHYD 449 449
FT CARBOHYD 462 462
FT CARBOHYD 528 528
FT CARBOHYD 642 642
FT CARBOHYD 694 694
FT CARBOHYD 857 857
FT CARBOHYD 894 894
FT CARBOHYD 973 973
FT CARBOHYD 1032 1032
FT CARBOHYD 1040 1040
FT VARIANT 433 433
FT VARIANT 524 524
FT VARIANT 972 972
FT VARIANT 1003 1003
FT VARIANT 1030 1030
FT VARIANT 1094 1094
FT SEQUENCE 1189 AA; 133609 MW; 60303C08A44ACD52 CRC64;

Query Match 53.1%; Score 544; DB 1; Length 1189;
Best Local Similarity 53.0%; Pred. No. 2,6e-41;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

1 GICARVDASFOPOGSLAFTAQRCPFTYMDVIVLIDGNSIYPSSEVQFLRLVGLFLDP 60
137 GMSCRVNSFRSKVPAALQRCQTYMDIVLIDGNSIYPSSEVQFLRLVGLFLDP 196
61 EGIQVGLVQGESPVHEMSLGFRTKEEVRAKNISREGRKETAQAIVACTEGFSG 120
197 GGIQVGVQYGGDVVHEHFLNDYRSKYDVAAASHIEQCGTETRAFIETFAKSEAPQK 256
121 SHGSEPAARLLIVVTGDSHGEELPAALKACEAGRVTRYGIIVLIGHLARQRPSSFL 180

```

```

Db 257 --GGKRAKKVMIVITGESHSDPLEKVIQQSERDNTRYAVAVLGYNRRGINPETFL 314
Qy 181 REIRTIASDPDPPEFFNV 198
Db 315 NEIKTIASDPDPKHFNV 332

RESULT 3
ID ITAI CHICK STANDARD; PRT; 285 AA.
AC Q90615;
DT 15-OUL-1998 (Rel. 36, Created)
DT 15-OUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
DE (Fragment).
GN ITCA1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Glizard;
RA MEDLINE=94357930; PubMed=7521332;
RX Kern A., Brissewitz R., Bank I., Marcantonio E.E.;
RT "The role of the I domain in ligand binding of the human integrin
alpha 1 beta 1."
RL J. Biol. Chem. 269:22811-22816(1994).
CC -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
E-R IN COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; U10114; AAA59067.1; -.
CC HSSP; P17301; IAOX.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00092; VWA; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; PARTIAL.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane.
FT NON_TER 1 1
FT DOMAIN 1 285
FT CARBOHYD 66 279
FT CARBOHYD 2 2
FT CARBOHYD 40 40
FT CARBOHYD 208 208
FT CARBOHYD 232 232
FT NON_TER 285 285
FT SEQUENCE 285 AA; 31503 MW; 1B05D3246CDSCA7E CRC64;

Query Match 46.1%; Score 473; DB 1; Length 285;
Best Local Similarity 45.5%; Pred. No. 1.2e-35;
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

1 GICARVDASFOPOGSLAFTAQRCPFTYMDVIVLIDGNSIYPSSEVQFLRLVGLFLDP 60

```

```

Db      36 GVCNSVSTFEETKAVAPSVQECCKTQDIDIVLDGNSIYPMBSVTAFLNLSLRNDIGP 95
QY      61 EOIQVGLVQGESPVHEWMSLGDFTREEVVRANKLSRREGRTKTAQAIVWACTGFSQ 120
Db      96 QOTQVIGVQGTQVHEFYNTYSTEEVMDALRLRQGTQMTALGDTAREBAFTE 155
QY      121 SHGCRPEARLLVVTVDGESHDEELPAALKACEAGRVTRYGIAVLGHYLRORDPSFL 180
Db      156 AHGARRGVQKVMIVITVDGESHNDYRQOEVIDKCEDENIQFAIILGTSRGNLSTEXEV 215
QY      181 REIRTIASDPDERFFNFV 198
Db      216 ERIKSIASEPTEKHPNFV 233

RESULT 4
ID      ITAI_HUMAN      STANDARD;      PRT; 1151 AA.
AC      P56139;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
GN      ITGA1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN      NCBI_TaxID=9606;
RX      MEDLINE=93155124; PubMed=8428973;
RA      Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT      Expression of native and truncated forms of the human integrin alpha
RT      1 subunit.
RL      J. Biol. Chem. 268:2989-2996(1993).
CC      -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC      COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC      E-R IN COLLAGEN.
CC      -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC      ASSOCIATES WITH BETA-1.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC      WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC      -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC      -1- DATABASE: NAME=PROV; NOTE=CD guide CD49a entry;
CC      WWW="http://www.ncbi.nlm.nih.gov/ProV/cd/cd49a.htm".
DR      HSSP; P17301; IAOX.
DR      Genew; HGNC:6134; ITGA1.
DR      MIM; 192968;
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam; PFO0092; vwa; 1.
DR      Pfam; PFO0357; Integrin_A; 1.
DR      Pfam; PFO1839; FG-GAP; 5.
DR      PRINTS; PR00453; VWFADOMAIN.
DR      SMART; SM00191; Int_alpha; 5.
DR      SMART; SM00327; VWFA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; VWFA; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Repeat; Calcium; Magnesium.
FT      DOMAIN 1 1113 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 1114 1136 POTENTIAL.
FT      DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).
FT      REPEAT 16 75 FG-GAP 1.
FT      REPEAT 7 75 FG-GAP 2.
FT      DOMAIN 147 360 VWFA.
FT      REPEAT 349 404 FG-GAP 3.
FT      REPEAT 405 457 FG-GAP 4.
FT      REPEAT 459 520 FG-GAP 5.

```

```

FT      REPEAT 540 599 FG-GAP 6.
FT      REPEAT 602 654 FG-GAP 7.
FT      CA_BIND 470 478 POTENTIAL.
FT      CA_BIND 552 560 POTENTIAL.
FT      CA_BIND 614 622 POTENTIAL.
FT      SITE 1139 1142 GEFKR MOTIF.
FT      DISULFID 54 64 BY SIMILARITY.
FT      DISULFID 660 669 BY SIMILARITY.
FT      DISULFID 675 728 BY SIMILARITY.
FT      DISULFID 780 786 BY SIMILARITY.
FT      DISULFID 850 858 BY SIMILARITY.
FT      DISULFID 1002 1034 BY SIMILARITY.
FT      DISULFID 1037 1044 BY SIMILARITY.
FT      CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1151 AA; 127837 MW; 6B3F3C1AAB52808 CRC64;

Query Match      45.6%; Score 467.5; DB 1; Length 1151;
Best Local Similarity 46.5%; Pred. No. 2,1e-34;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

QY      1 GICARVDSFQPGSLAPTAQRCPYMDVIVVLGNSIYPMBSVQFTRRLVGLFDP 60
Db      118 GICSVSPTFPVQVNSIAP-VOECSTQDIDIVLDGNSIYPMBSVTAFLNLSLRNDIGP 176
QY      61 EOIQVGLVQGESPVHEWMSLGDFTREEVVRANKLSRREGRTKTAQAIVWACTGFSQ 120
Db      177 KOTQVIGVQGTQVHEFYNTYSTEEVMDALRLRQGTQMTALGDTAREBAFTE 236
QY      121 SHGCRPEARLLVVTVDGESHDEELPAALKACEAGRVTRYGIAVLGHYLRORDPSFL 180
Db      237 ARGARRGVQKVMIVITVDGESHNDYRQOEVIDKCEDENIQFAIILGTSRGNLSTEXEV 296
QY      181 REIRTIASDPDERFFNFV 198
Db      297 ERIKSIASEPTEKHPNFV 314

RESULT 5
ID      ITAI_RAT      STANDARD;      PRT; 1180 AA.
AC      P18614;
DT      01-NOV-1990 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE      (CD49a).
GN      ITGA1.
OS      Rattus norvegicus (Rat).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxId=10116;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=90338125; PubMed=2380249;
 CC Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A.,
 CC Esch F., Carbonetto S., Reichardt L.F.;
 CC "Molecular cloning of the rat integrin alpha 1-subunit: a receptor
 CC for laminin and collagen.";
 CC J. Cell Biol. 111:709-720(1990).
 CC [2]
 CC X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
 CC MEDLINE=99313197; PubMed=10386626;
 CC Nole M., Pepinsky R.B., Vanyamov S.Y., Koteliansky V.,
 CC Gotwals P.J., Kapusas M.;
 CC "Crystal structure of the alphabeta integrin I-domain: insights into
 CC integrin I-domain function.";
 CC FEBS Lett. 452:379-385(1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
 CC COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
 CC E-R IN COLLAGEN.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL, X52140; CAA36384.1; -.
 CC DR PIR, A35854; A35854.
 CC DR HSSP, P17301; IAOX.
 CC DR InterPro, IPR000413; Integrin_alpha.
 CC DR InterPro, IPR002035; WVF_A.
 CC DR Pfam, PF00092; vwa; 1.
 CC DR Pfam, PF00357; integrin_A; 1.
 CC DR Pfam, PF01839; FG-GAP_5.
 CC DR PRINTS, PR00453; VWFADOMAIN.
 CC DR SMART, SM00191; Int_alpha; 5.
 CC DR SMART, PR00453; VWFADOMAIN.
 CC DR PROSITE, PS00242; INTEGRIN_ALPHA; 1.
 CC DR PROSITE, PS50234; VWFA; 1.
 CC DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 CC Signal; Repeat; Calcium; Magnesium.
 CC KM SIGNAL 1 28
 CC FT CHAIN 29 1180 INTEGRIN ALPHA-1.
 CC FT DOMAIN 29 1142 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 1143 1165 POTENTIAL.
 CC FT DOMAIN 1166 1180 CYTOPLASMIC (POTENTIAL).
 CC FT REPEAT 44 103 FG-GAP 1.
 CC FT REPEAT 175 388 FG-GAP 2.
 CC FT DOMAIN 377 432 VWFA.
 CC FT REPEAT 433 484 FG-GAP 3.
 CC FT REPEAT 485 565 FG-GAP 4.
 CC FT REPEAT 567 626 FG-GAP 5.
 CC FT REPEAT 629 681 FG-GAP 6.
 CC FT CA_BIND 497 505 FG-GAP 7.
 CC FT CA_BIND 579 649 POTENTIAL.
 CC FT CA_BIND 641 649 POTENTIAL.
 CC FT SITE 1168 1172 GFPK MOTIF.
 CC FT DISULFID 82 92 BY SIMILARITY.
 CC FT DISULFID 687 696 BY SIMILARITY.

FT DISULFID 702 755 BY SIMILARITY.
 FT DISULFID 807 813 BY SIMILARITY.
 FT DISULFID 877 885 BY SIMILARITY.
 FT DISULFID 1029 1062 BY SIMILARITY.
 FT DISULFID 1066 1073 BY SIMILARITY.
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 698 698 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 779 779 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 820 820 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 907 907 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1180 AA; 130808 MW; 8E5DA2BE02362BE4 CRC64;

Query Match 45.4%; Score 465.5; DB 1; Length 1180;
 Best Local Similarity 46.5%; Pred. No. 3.3e-34;
 Matches 92; Conservative 35; Mismatches 70; Indels 1; Gaps 1;

QY 1 GICARVDASFPQGSLATACRCPTMDVYVLDGNSITYWSEVQTRLRVGLKFLIDP 60
 DB 146 GICSDVSPFPQVNVSPAP-VOECSTQDIVIVLDGNSIYWPESVIAFLNLDKMDIGP 204
 QY 61 BOIQVGLVQGESPPVHEMSLDGPFRTKEVVAANKLSRREGRETKTAQAIWVACGEPSF 120
 DB 205 KOTQVIGVQSEBNTHEENLKYSSTEBVLAANKIGQGLQMTALGIDTPARKEATTE 264
 QY 121 SHGGRPEARLVVVTGDESHDGEELPALACACAGRTYGVIVLGHVLRORPSSFL 180
 DB 265 ARGARGVKXKVMVITVTDGSHDNRLKQVIOQDCEDENIGRSIALIGHYNGNSTEKFFV 324
 QY 181 REIRTIASDPDERFFENV 198
 DB 325 EIKSIASEPTEKHFENV 342

RESULT 6
 ID ITA2 MOUSE STANDARD; PRT, 1178 AA.
 AC Q62469; Q62163;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)
 DB (Collagen receptor) (VLA-2 alpha chain) (CD49b).
 GN ITGA2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxId=10090;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 CC MEDLINE=94363406; PubMed=8081889;
 CC Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,
 CC Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.;
 CC "The mouse VLA-2 homologue supports collagen and laminin adhesion but

RT not virus binding." ;
 RL Cell Adhes. Commun. 2:131-143 (1994).
 RN [2]
 RP SEQUENCE OF 450-1178 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=94355691; PubMed=7521231;
 RA Wu J.E., Santoro S.A.;
 RT "Complex patterns of expression suggest extensive roles for the alpha
 2 beta 1 integrin in murine development." ;
 RL Dev. Dyn. 199:292-314 (1994).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOLOGOUS FOR A NULL MUTATION
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z29987; CA82877.1; -;
 DR EMBL: X75427; CA83178.1; -;
 DR HSSP: P17301; IAOX.
 DR MGD: MGI:96600; ITC2.
 DR InterPro: IPR000413; Integrin_alpha.
 DR InterPro: IPR002035; VFMA_A.
 DR Pfam: PF00092; vma; 1.
 DR Pfam: PF00357; Integrin_A; 1.
 DR Pfam: PF01839; FG-GAP; 5.
 DR PRINTS: PR00453; VFMADOMAIN.
 DR SMART: SM00191; Int_alpha; 5.
 DR SMART: SM00327; VMA_1.
 DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE: PS0234; VFMA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Platelet; Signal; Repeat; Calcium; Magnesium.
 FT SIGNAL 1 26
 FT CHAIN 27 1178
 FT DOMAIN 27 1129
 FT TRANSMEM 1130 1151
 FT DOMAIN 1152 1178
 FT REPEAT 42 100
 FT REPEAT ? ?
 FT DOMAIN 185 375
 FT REPEAT ? ?
 FT REPEAT 431 483
 FT REPEAT 485 546
 FT REPEAT 548 607
 FT REPEAT 612 664
 FT CA_BIND 496 568
 FT CA_BIND 560 568
 FT CA_BIND 624 632
 FT SITE 480 482
 FT SITE 1154 1158
 FT DISULFID 80 89
 FT DISULFID 677 734
 FT DISULFID 786 792
 FT DISULFID 862 873
 FT DISULFID 1016 1047

FT DISULFID 1052 1057 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 696 696 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1054 1054 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B3C0240F465 CRC64;
 Query Match 44.3%; Score 454; DB 1; Length 1178;
 Best Local Similarity 46.0%; Pred. No. 3.6e-33;
 Matches 91; Conservative 33; Mismatches 74; Indels 0; Gaps 0;
 QY 1 GICARVDASFOGSLAFTACRPIYMDVYVLDGNSIYFWSVQTLRLVGLFLDP 60
 DB 144 GICSDVSPDFQFLTFSFAVQACPSLDVDDVVCDESNISYFWEAVKFLVRFVGLDIGP 203
 QY 61 EGIQVGIYOGESPVHEMSLDGPFRTKEVRAAKNLSRREGRETKQAQIIVACTEGFSQ 120
 DB 204 KKTQVALIQYANEPRIINLNDPFTKEDVQATSETRHGGLDITVFAIFARDYASQ 263
 QY 121 SHGGRPEARLLVVTGDESHDGEELPAALACENAGRYGIAVLGHYLRORDPSSFL 180
 DB 264 TSGGRPGATKVVVVTGDESHDGSGLKTVIQCNDDLELRGIAVLGYLNNAIDTKLI 323
 QY 181 REIRTIASDPDRFFPNV 198
 DB 324 KEIKAIASPTPERYFENV 341
 RESULT 7
 TR2_BOVIN
 ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
 AC P53710;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
 GN ITGA2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94193647; PubMed=7511592;
 RA Kamata T., Puzon W., Takada Y.;
 RT "Identification of putative ligand binding sites within I domain of
 RL Integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
 RJ J. Biol. Chem. 269:9659-9663 (1994).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
 CC CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE
 CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X70793; CAA50064.1; -
 DR EMBL; X70792; CAA50063.1; -
 DR EMBL; X66138; CAA46928.1; -
 DR PIR; S22916; S22916.
 DR PDB; 1B9P; 25-FEB-99.
 DR PDB; 1B9Q; 25-FEB-99.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR003962; FNIII_repeat.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; WVF_A.
 DR Pfam; PF00041; fn3; 8.
 DR Pfam; PF00092; wva; 2.
 DR Pfam; PF01391; Collagen; 4.
 DR Pfam; PF02310; TSPN; 1.
 DR PRINTS; PR00014; FNTYPEIII.
 DR PRINTS; PR00453; WVFADOMAIN.
 DR SMART; SM00060; FN3; 7.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; WVA; 2.
 DR PROSITE; PS50234; WVFA; 2.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KM Cell adhesion; Collagen; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 28
 FT CHAIN 1 1888
 FT DOMAIN 158 330 WVFA 1.
 FT DOMAIN 441 529 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 531 622 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 625 728 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 739 821 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 829 922 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 952 1010 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 1042 1215 WVFA 2.
 FT DOMAIN 1227 1468 NONHELICAL REGION (NC4).
 FT DOMAIN 1469 1620 TRIPLE-HELICAL REGION 1.
 FT DOMAIN 1664 1786 TRIPLE-HELICAL REGION 2.
 FT DOMAIN 721 733 POLY-THR.
 FT CAROHXD 138 138 N-LINKED (GICNAC. . .) (POTENTIAL).
 FT CAROHXD 1398 1398 N-LINKED (GICNAC. . .) (POTENTIAL).
 FT SITE 1489 1491 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1617 1619 CELL ATTACHMENT SITE (POTENTIAL).
 SQ SEQUENCE 1888 AA; 202666 MW; 39915BB9F46DD873 CRC64;

Query Match 22.2% Score 227.5; DB 1; Length 1888;
 Best Local Similarity 31.7%; Pred. No. 2e-12; Indels 17; Gaps 6;
 Matches 60; Conservative 37; Mismatches 75;

QY 12 PGSLAPTAQRCPTVDVIVLVDGNSI--YMWSEVQFLRLVGLFLFDPEQIOGVAVO 69
 DB 144 PEGNFTFC-KTPALADIIVLVDGWSIGRFNRLVRLLENLVAFNVGSEKTVGLAQ 201
 QY 70 YGESPVHWSLGDFTKEEVPAKKLSRREGRETKTAQAIWACTEGFSQSHGRPEEA 129
 DB 202 YSGDRIEWHNLAVYGTDAVAVRLPYK-GGNLTGTALTYILENSFKPRAGARPGVS 260
 QY 130 RLIVVTVDSHGEELPAALAKACEAGRTRYGIVLGHYLRORDPSFLREIRTIASD 189
 DB 261 KGIILITDKSD-DVIRPAKRLRDG-IELPAIGV-----KNADINELKEINASE 308
 QY 190 PDREFFNV 198
 DB 309 PDSTHYNV 317

RESULT 10
 CAIC_CHICK

ID CAIC_CHICK STANDARD; PRT; 3124 AA.
 AC P13944; Q04509;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor (Fibrochimerin).
 GN COL12A1.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OK NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=White Leghorn;
 RX MEDLINE=92011862; PubMed=1918137;
 RA Yamagata M., Yamada K.M., Yamada S.S., Shinomura T., Tanaka H.,
 RA Nishida Y., Obara M., Kimata K.;
 RT "The complete primary structure of type XII collagen shows a chimeric
 RT molecule with reiterated fibronectin type III motifs, von Willebrand
 RT factor A motifs, a domain homologous to a noncollagenous region of
 RT type IX collagen, and short collagenous domains with an Arg-Gly-Asp
 RT site."
 RL J. Cell Biol. 115:209-221 (1991).
 RN [2]
 RP SEQUENCE OF 2456-3124 FROM N.A., AND SEQUENCE OF 2772-2794 AND
 RP 2846-2873.
 RX MEDLINE=90062079; PubMed=2584192;
 RA Gordon M.K., Gerecke D.R., Dublet B., van der Rest M., Olsen B.R.;
 RT "Type XII collagen: A large multidomain molecule with partial
 RT homology to type IX collagen."
 RL J. Biol. Chem. 264:19772-19778 (1989).
 RN [3]
 RP SEQUENCE OF 2960-3076 FROM N.A.
 RX MEDLINE=87317590; PubMed=3476925;
 RA Gordon M.K., Gerecke D.R., Olsen B.R.;
 RT "Type XII collagen: distinct extracellular matrix component
 RT discovered by cDNA cloning."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6040-6044 (1987).
 RN [4]
 RP SEQUENCE OF 1-1283 FROM N.A. (SHORT FORM), AND ALTERNATIVE SPLICING.
 RC TISSUE=Embryo;
 RX MEDLINE=93042014; PubMed=1420368;
 RA Trueb J., Trueb B.;
 RT "The two splice variants of collagen XII share a common 5' end."
 RL Biochim. Biophys. Acta 1171:97-98 (1992).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=95370352; PubMed=7642694;
 RA Koch M., Bohrmann B., Mathison M., Hagios C., Trueb B., Chiquet M.;
 RT "Large and small splice variants of collagen XII: differential
 RT expression and ligand binding."
 RL J. Cell Biol. 130:1005-1014 (1995).
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
 CC CONTAINING FIBRILS, THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
 CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
 CC LOCALIZED IN THE PERITRIBULAR MATRIX.
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
 CC NONTRIPLE-HELICAL SEQUENCES.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE FINAL TISSUE
 CC FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OF EITHER THE LONGER
 CC OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT
 CC ISOFORM CHAINS. ONLY THE LONG VARIANT IS A PROTEOLYTIC. THE LARGE
 CC ISOFORM HAS MORE RESTRICTED EXPRESSION IN EMBRYONIC TISSUE THAN
 CC THE SMALL.
 CC -1- TISSUE SPECIFICITY: TYPE XII COLLAGEN IS PRESENT IN TENDONS,
 CC LIGAMENTS, PERICHOONDRIUM, AND PERIOSTEUM.
 CC -1- TISSUES CONTAINING TYPE I COLLAGEN.
 CC -1- DOMAIN: THIS SEQUENCE DEFINES FIVE DISTINCT DOMAINS. TWO TRIPLE-
 CC HELICAL DOMAINS (COL1 AND COL2) AND THREE NONTRIPLE-HELICAL
 CC DOMAINS (NC1, NC2, AND NC3).
 CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT

CC EACH END.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAIN.
 CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -1- SIMILARITY: CONTAINS 4 WFPA DOMAINS.
 CC -1- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D00824; BAA00701.1; -
 CC EMBL; X61024; CAA43358.1; -
 CC EMBL; M17375; AAA48718.1; -
 CC EMBL; J05137; AAA48635.1; -
 CC EMBL; X67327; CAA47744.1; -
 CC PIR; A28037; A28037.
 CC PIR; A34485; A34485.
 CC HSSP; P20701; ILFA.
 CC InterPro; IPR000087; Collagen.
 CC InterPro; IPR003961; FN III.
 CC InterPro; IPR003962; FNIII_repeat.
 CC InterPro; IPR003129; TSPN.
 CC InterPro; IPR002035; WVF_A.
 CC Pfam; PF00041; fn3; 17.
 CC Pfam; PF00092; vwa; 4.
 CC Pfam; PF01391; Collagen; 4.
 CC Pfam; PF02210; TSPN; 1.
 CC PRINTS; PR00014; ENTPEP11.
 CC PRINTS; PR00453; WVFADOMAIN.
 CC SMART; SM00060; FN3; 17.
 CC SMART; SM00327; TSPN; 1.
 CC PROSITE; PS50234; WVFA; 4.
 CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.
 CC
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 3124
 CC FT DOMAIN 25 114
 CC FT DOMAIN 139 311
 CC FT DOMAIN 332 425
 CC FT DOMAIN 439 615
 CC FT DOMAIN 629 720
 CC FT DOMAIN 721 811
 CC FT DOMAIN 812 904
 CC FT DOMAIN 905 998
 CC FT DOMAIN 999 1085
 CC FT DOMAIN 1086 1178
 CC FT DOMAIN 1178 1371
 CC FT DOMAIN 1199 1371
 CC FT DOMAIN 1384 1473
 CC FT DOMAIN 1474 1565
 CC FT DOMAIN 1566 1654
 CC FT DOMAIN 1655 1755
 CC FT DOMAIN 1756 1846
 CC FT DOMAIN 1847 1936
 CC FT DOMAIN 1937 2027
 CC FT DOMAIN 2028 2118
 CC FT DOMAIN 2119 2206
 CC FT DOMAIN 2207 2295
 CC FT DOMAIN 2327 2500
 CC FT DOMAIN 2455 2750
 CC FT DOMAIN 2751 2902
 CC FT DOMAIN 2903 2945
 CC FT DOMAIN 2946 3048
 CC FT DOMAIN 3049 3124

FT DOMAIN 3086 3096 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 3111 3123 ARG/LYS-RICH (BASIC).
 FT BINDING 797 797 TO CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 890 890 TO CHONDROITIN SULFATE (POTENTIAL).
 FT BINDING 981 981 TO CHONDROITIN SULFATE (POTENTIAL).
 FT SITE 2438 2440 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2899 2901 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1006 1006 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1032 1032 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1512 1512 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1767 1767 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2210 2210 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2273 2273 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2532 2532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2683 2683 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 25 1188 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 1258 1258 T -> S (IN REF. 4).
 FT CONFLICT 1264 1264 D -> E (IN REF. 4).
 FT CONFLICT 2759 2759 P -> A (IN REF. 2).
 FT CONFLICT 2803 2803 L -> F (IN REF. 2).
 FT CONFLICT 2977 2977 V -> F (IN REF. 2).
 FT CONFLICT 3075 3076 QP -> AG (IN REF. 3).
 SQ SEQUENCE 3124 AA; 340578 MW; 094285AFET7346CF CRC64;
 Query Match 22.2%; Score 227.5; DB 1; Length 3124;
 Best Local Similarity 34.1%; Pred. No. 3.7e-12;
 Matches 59; Conservative 32; Mismatches 67; Indels 15; Gaps 5;
 QY 28 DVIVIVLDGNSI--YPMSEVOTFLRLVYKLFIDPEQIQVGLVQGESPVHEWSIGDFT 85
 DB 1199 DIVLVDSWMSIGRNFVFNFRISRIVEFDIGDKYQIGLAQSGSPRIEMNLNAVRT 1258
 QY 86 KEAVYRAKUNLSRREGRETQAQAIMVACTGSEFSGHSGRREARLLVVTVDGESHGDEE 145
 DB 1259 KEALDAVYTNLPYK-GGNTLLTGMAIDFLKNNFKDEALRPARRIGVLITDGSKDDVV 1317
 QY 146 LPALKACEAGVTRYGIAVLGHYLRORPDSFIREIRTTASDPDERFFNV 198
 DB 1318 TPSSRLRDEG--VELYALGI-----KNADN----ELKQIATDPDDIHAINV 1358
 RESULT 11
 CAVA_CHICK STANDARD; PRT; 493 AA.
 AC P05059;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cartilage matrix protein precursor (Matrilin-1).
 GN MATN1 OR CMP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-43.
 RC STRAIN=White Leghorn;
 RC MEDLINE=89255246; PubMed=2542265;
 RA Kiss I., Deak F., Holloway R.G., Jr., Delius H., Mebust K.A.,
 RA Fritberger E., Argaves W.S., Tsomis P.A., Winterbottom N.,
 RA Goelink P.F.;
 RT "Structure of the gene for cartilage matrix protein, a modular
 RT protein of the extracellular matrix. Exon/intron organization,
 RT unusual splice sites, and relation to alpha chains of beta 2
 RT integrins, von Willebrand factor, complement factors B and C2, and
 RT epidermal growth factor.";
 RL J. Biol. Chem. 264:8126-8134 (1989).
 RN [2]
 RP SEQUENCE OF 78-493 FROM N.A.
 RX MEDLINE=87092429; PubMed=3025875;

RA Argraeve W.S., Deak F., Sparks K.J., Kiss I., Goetinck P.F.,
 RT "Structural features of cartilage matrix protein deduced from cDNA."
 RL Proc. Natl. Acad. Sci. U.S.A. 84:464-468(1987).
 RN [3]
 RP STRUCTURE BY NMR OF 447-493.
 RX MEDLINE=97406913; PubMed=9260286;
 RA Wilschack R., Kammerer R.A., Dames S.A., Schulthess T.,
 RA Blomere M.J., Engel J., Alexandrescu A.T.,
 RT "Heteronuclear NMR assignments and secondary structure of the coiled
 RT coil trimerization domain from cartilage matrix protein in oxidized
 RT and reduced forms."
 RL Protein Sci. 6:1734-1745(1997).
 CC -1- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE
 CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO
 CC COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 VWFA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X12353; CAA30915.1; -
 DR EMBL; X12346; CAA30915.1; JOINED.
 DR EMBL; X12347; CAA30915.1; JOINED.
 DR EMBL; X12348; CAA30915.1; JOINED.
 DR EMBL; X12349; CAA30915.1; JOINED.
 DR EMBL; X12350; CAA30915.1; JOINED.
 DR EMBL; X12351; CAA30915.1; JOINED.
 DR EMBL; X12352; CAA30915.1; JOINED.
 DR EMBL; M14792; AAA48695.1; -
 DR EMBL; M97497; AAC1872.1; -
 DR PIR; A33809; A33809.
 DR PIR; A26364; A26364.
 DR PDB; 1A05; 11-FEB-98.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00098; EGF; 1.
 DR Pfam; PF00092; VWF; 2.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00181; EGF; 1.
 DR SMART; SM00327; VWF; 2.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50234; VWF; 2.
 KW EGF-like domain; Signal; Glycoprotein; Cartilage; Repeat;
 KW Coiled coil; 3D-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 493 CARTILAGE MATRIX PROTEIN.
 FT DOMAIN 24 220 VWFA 1.
 FT DOMAIN 221 261 EGF-LIKE.
 FT DOMAIN 262 450 VWFA 2.
 FT DOMAIN 462 482
 FT CAROHD 74 74 COILED COIL.
 FT DISULFID 33 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 225 236 POTENTIAL.
 FT DISULFID 232 245 BY SIMILARITY.
 FT DISULFID 247 260 BY SIMILARITY.
 FT DISULFID 263 449 POTENTIAL.
 SQ SEQUENCE 493 AA; 54034 MW; E4D3DACFA2B96A4 CRC64;

Query Match 20.9%; Score 214; DB 1; Length 493;
 Best Local Similarity 33.3%; Pred. No. 6; Se-12;
 Matches 58; Conservative 35; Mismatches 63; Indels 18; Gaps 6;

QY 27 MUVVVLDSNSIYP--MSEVOTFLRLVYGLKFIPIDEQIQGLVYQSGSPVHMSIGDPR 84
 DB 271 LDLVFLIDSKSVRPENFELVKKFIQIVIESLVESEKQAVGLVQYSSSVROEFPFLGQPK 330

QY 85 TKEVYRAAKNLRSREGRTYTAQAIWVACTGSPQSHGHPREARLLVVTGESHGDE 144
 DB 331 NKKDIIKAIVKKAYWE-KGTMTGALKXLYLVSSFSFIANGARPGVKGIVFDGSSQD-Y 388
 QY 145 ELPAALKACENG-RYTRYGIAVLGHYLRQDPSSFLREITIASDP-DEPFF 196
 DB 389 ITDAKKKAKDGFRRFVAVGV-----NAVEDELREIASEPVAEHYFY 430
 RESULT 12
 ID CALC_MOUSE STANDARD; PRT; 3119 AA.
 AC Q60847; P70322;
 DT 15-JUN-1998 (Rel. 36, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XII) chain precursor.
 GN COL12A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-1 AND
 RP XI1B-1).
 RP STRAIN=Swiss Webster, and C57BL/6J; TISSUE=Skin;
 RX MEDLINE=96170761; PubMed=8601036;
 RA Boehme K., Li Y., Oh P.S., Olsen B.R.;
 RT "Primary structure of the long and short splice variants of mouse
 RT collagen XII and their tissue-specific expression during embryonic
 RT development."
 RL Dev. Dyn. 204:432-445(1995).
 RN [2]
 RP PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING (ISOFORMS XI1A-2
 RP AND XI1B-2).
 RP STRAIN=C57BL/6J; TISSUE=Skin fibroblast;
 RX MEDLINE=99348349; PubMed=10419532;
 RA Kanita A.M., Reichenberger E., Baur S.T., Karimbux N.Y., Taylor R.W.,
 RA Olsen B.R., Nishimura I.;
 RT "Structural variation of type XII collagen at its carboxyl-terminal
 RT NC1 domain generated by tissue-specific alternative splicing."
 RL J. Biol. Chem. 274:22053-22059(1999).
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
 CC CONTAINING FIBRILS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE
 CC SURFACE OF THE FIBRILS, AND THE COL2 AND NC3 DOMAINS MAY BE
 CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
 CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; XI1A-1 (SHOWN HERE), XI1A-
 CC 2/ERR#, XI1B-1 AND XI1B-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC THE FINAL TISSUE FORM OF COLLAGEN XII MAY CONTAIN HOMOTRIMERS OR
 CC ANY COMBINATION OF THE VARIOUS ISOFORMS.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TENDONS, PERICHONDRUM,
 CC SKIN, CORNEA, SCLERA, BLOOD VESSELS, AND PERIOSTEUM.
 CC -1- DEVELOPMENTAL STAGE: THE LONG NC3 XI1A ISOFORMS ARE PREDOMINANT AT
 CC EARLY STAGES (ED7 AND 11); AT LATER STAGES OF DEVELOPMENT (ED15
 CC AND 17) THE SHORT NC3 XI1B FORMS BECOME THE MAJOR FORMS. AS THE
 CC SHORT NC3 FORMS BECOME THE MAJOR PRODUCT, THE LONG SPLICED VARIANT
 CC CONTINUES TO BE EXPRESSED IN SEVERAL TISSUES, EVEN AFTER BIRTH.
 CC THE LONG NC1 ISOFORMS, XI1A-1 AND XI1B-1, PEAK IN 15-DAY OLD
 CC EMBRYOS AND DECREASE IN 17-DAY OLD ONES. THE EXPRESSION OF THE
 CC SHORT NC1 FORM XI1B-2 REMAINS CONSTANT THROUGHOUT LATE STAGES OF
 CC EMBRYONIC DEVELOPMENT (ED15 AND ED17).
 CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
 CC EACH END (BY SIMILARITY).
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
 CC SIMILARITY).
 CC -1- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY
 CC ISOFORM XI1A-2 IS A PROTEOGLYCAN.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.

FT	MOD_RES	2966	2966		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2969	2969		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2972	2972		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2984	2984		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3001	3001		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3004	3004		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3015	3015		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3024	3024		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3027	3027		HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3030	3030		HYDROXYLATION (BY SIMILARITY).
FT	DONAIN	869	872		POLY-THR.
FT	CARBOHYD	704	704		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1769	1769		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2212	2212		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2533	2533		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2684	2684		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	25	1190		MISSING (IN ISOFORM XIIB-1 AND ISOFORM XIIB-2).
FT	VARSPLIC	3062	3064		EPY -> GSG (IN ISOFORM XIIA-2 AND ISOFORM XIIB-2).
FT	VARSPLIC	3065	3119		MISSING (IN ISOFORM XIIA-2 AND ISOFORM XIIB-2).
SQ	SEQUENCE	3119 AA;	340239 MW;	981F999CG6A8J251 CXC64;	
 Query Match					
Best Local Similarity		20.8%;	Score 213.5;	DB 1;	Length 3119;
Matches		55;	Conservative 33;	Mismatches 70;	Indels 15; Gaps 4;
Oy	28	DVIVLVGSSNSI--YPWSEVOTFLRLVGLFTIDPEIOYGLVOYGESPHEWSIGDFRT	85		
Db	1203	DIVLVGVSSNSIGRANFRTRVSFSRIVEVPFLCPKVVQALAOISSDPTEMOIANHRD	1262		
Oy	86	KEEVARAAKNLSRRREGRETYTAQAIVNACTEGESQSHGSRPEARALLVVTDSGSHDGE	145		
Db	1263	KKSLLQAVAMLPNK-GGNTLTGMALNPFIPOGSFQTQAGMPRAKKISGVLITDGSKPDVE	1321		
Oy	146	LPAALXACEAGRVRVCIALVGHILRRQRPPSSLREIRITASDPDRFFENV	198		
Db	1322	APSFKLKDEGEVELFAISI-----KNABEV-----ELKMIAIDPDPTAHNV	1362		
 RESULT 13					
ID	CAIC_HUMAN	STANDARD;	PRT;	3063	AA.
AC	Q99715;	Q99716;			
DT	15-JUL-1998	(Rel. 36,	Created)		
DT	15-JUL-1998	(Rel. 36,	Last sequence update)		
DT	15-JUN-2002	(Rel. 41,	Last annotation update)		
DE	Collagen alpha 1(XII) chain precursor.				
GN	COL12A1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxId=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (LONG AND SHORT FORM), AND PARTIAL SEQUENCE.				
RX	MEDLINE=972885521; PubMed=9143499;				
RA	Gerdecke D.R., Olson P.F., Koch W., Knoll J.H.M., Taylor R.,				
RA	Hudson D.L., Champilaud M.-F., Olsen B.R., Bugesen R.E.;				
RT	"Complete primary structure of two splice variants of collagen XII,				
RT	and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX)				
RT	collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human				
RT	chromosome 6p12-q13."				
RL	Genomics 41:236-242(1997).				
CC	-I- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-				
CC	CONTAINING FIBRILLS. THE COL1 DOMAIN COULD BE ASSOCIATED WITH THE				
CC	SURFACE OF THE FIBRILLS, AND THE COL2 AND NC3 DOMAINS MAY BE				
CC	LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY).				
CC	-I- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF				
CC	NONTRIPLE-HELICAL SEQUENCES.				
CC	-I- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A				
CC	SHORT FORM, ARE PRODIED BY ALTERNATIVE SPLICING. THE FINAL TISSUE				
CC	FORM OF COLLAGEN XII MAY CONTAIN HOMOMERISERS OF EITHER THE LONGER				

OR THE SHORTER ISOFORM OR ANY COMBINATION OF LONG AND SHORT ISOFORM CHAINS.

- TISSUE SPECIFICITY: FOUND IN COLLAGEN 1-CONTAINING TISSUES: BOTH SHORT AND LONG ISOFORMS APPEAR IN ANION, CHORION, SKELETAL MUSCLE, SMALL INTESTINE, AND IN CELL CULTURE OF DERMAL FIBROBLASTS, KERATINOCYTES, AND ENDOTHELIAL CELLS. ONLY THE SHORT ISOFORM IS FOUND IN LUNG, PLACENTA, KIDNEY, AND A SQUAMOUS CELL CARCINOMA CELL LINE.

- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT EACH END (BY SIMILARITY).

- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY SIMILARITY).

- PTM: GLYCOSYLATED, CHONDROITIN-SULFATED (BY SIMILARITY). ONLY THE LONG VARIANT IS A PROTEOGLYCAN.

- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.

- SIMILARITY: CONTAINS 4 WFPA DOMAINS.

- SIMILARITY: CONTAINS 18 FIBRONECTIN TYPE III-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U73778; AAC51244.1; -
EMBL: U73779; AAD0483.1; -
HSSP: P02751; 1TTF.
Genew: HGNC:2188; COL12A1.
MIM: 120320; -
InterPro: IPR000087; Collagen.
InterPro: IPR003961; FN_III.
InterPro: IPR003962; FNIII_repeat.
InterPro: IPR003129; TSPN.
InterPro: IPR002035; WMF_A.
Pfam: PF00041; fn3; 18.
Pfam: PF00092; wfa; 4.
Pfam: PF01391; Collagen; 4.
Pfam: PF02210; TSPN; 1.
PRINTS: PR00014; FNTPETII.
PRINTS: PR00453; WMPADOMAIN.
SMART: SM00060; FN3; 17.
SMART: SM00210; TSPN; 1.
SMART: SM00327; WMA; 4.
PROSITE: PS50234; WFPA; 4.
Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Signal; Glycoprotein; Alternative splicing.

SIGNAL 1 25 3063 COLLAGEN ALPHA 1 (XII) CHAIN.
FT CHAIN 24 24
FT DOMAIN 25 114
FT DOMAIN 140 316 WFPA 1.
FT DOMAIN 333 426 FIBRONECTIN TYPE-III 1.
FT DOMAIN 440 616 WFPA 2.
FT DOMAIN 630 721 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 812 FIBRONECTIN TYPE-III 4.
FT DOMAIN 813 903 FIBRONECTIN TYPE-III 5.
FT DOMAIN 904 998 FIBRONECTIN TYPE-III 6.
FT DOMAIN 999 1085 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1086 1178 FIBRONECTIN TYPE-III 8.
FT DOMAIN 1199 1371 WFPA 3.
FT DOMAIN 1384 1473 FIBRONECTIN TYPE-III 9.
FT DOMAIN 1474 1564 FIBRONECTIN TYPE-III 10.
FT DOMAIN 1565 1652 FIBRONECTIN TYPE-III 11.
FT DOMAIN 1654 1751 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1752 1842 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1843 1932 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1933 2023 FIBRONECTIN TYPE-III 15.
FT DOMAIN 2024 2114 FIBRONECTIN TYPE-III 16.
FT DOMAIN 2115 2202 FIBRONECTIN TYPE-III 17.
FT DOMAIN 2203 2291 FIBRONECTIN TYPE-III 18.

FT	DOMAIN	2323	2496	WFPA 4.
FT	DOMAIN	2451	2746	NONHELICAL REGION (NC3).
FT	DOMAIN	2747	2898	TRIPLE-HELICAL REGION (COL2) WITH 1 IMPERFECTION.
FT	DOMAIN	2899	2941	NONHELICAL REGION (NC2).
FT	DOMAIN	2942	3044	TRIPLE-HELICAL REGION (COL1) WITH 2 IMPERFECTIONS.
FT	DOMAIN	3045	3063	NONHELICAL REGION (NC1).
FT	BINDING	798	798	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	889	889	TO CHONDROITIN SULFATE (POTENTIAL).
FT	BINDING	981	981	TO CHONDROITIN SULFATE (POTENTIAL).
FT	SITE	862	864	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2779	2781	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	2895	2897	CELL ATTACHMENT SITE (POTENTIAL).
FT	MOD_RES	2944	2944	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2947	2947	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2950	2950	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2952	2959	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2965	2965	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2968	2968	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2971	2971	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	2983	2983	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3000	3000	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3003	3003	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3014	3014	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3023	3023	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3026	3026	HYDROXYLATION (BY SIMILARITY).
FT	MOD_RES	3029	3029	HYDROXYLATION (BY SIMILARITY).
FT	CARBOHYD	700	700	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	1763	1763	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	2206	2206	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	2528	2528	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD	2679	2679	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	VARSPLIC	25	1188	MISSING (IN SHORT ISOFORM).
SEQ	SEQUENCE	3063 AA;	333189 MW;	75FPA78FPA848293 CRC64;

Query Match 20.4%; Score 209.5; DB 1; Length 3063;
Best Local Similarity 31.8%; Pred. No. 1.6e-10;
Matches 55; Conservative 33; Mismatches 70; Indels 15; Gaps 4;

QY	28	DVIVLIDGDSNST--YPMSEVQTFRLRVGKLFIDPEQOVGLVYGESPNHMSIGDRT	85
DB	1199	DIVLIVDGSWSTIGRANFRTVTSFISRIIVEVDIGKRVQVLAQSGDPRTWQVNAHRD	1258
QY	86	KEEVVRAKNLSRRGRSTKTAQAIMVACTGFSQSHGGRPREARLVVVDGESHGDEE	145
DB	1259	KKSLLQAVANLPYK-GGNTLTGMALNFRQONFRQAGMRPRARKIGVLTIDGKSQDDVE	1317
QY	146	LPALKACEAGRVTRYGIAVLGHVLRQRDPSSFLREIRRTASDPERFFENV	198
DB	1318	APSKLXKDEGVLELPAIGI-----KNADEV---ELKMLATDDDDHNVV	1358

RESULT 14

ID	CAMA HUMAN	STANDARD;	PRT;	496 AA.
AC	P21941;			
DT	01-MAY-1991 (Rel. 18, Created)			
DT	01-MAY-1991 (Rel. 18, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cartilage matrix protein precursor (Matrilin-1).			
GN	MATN1 OR CRTM OR CMP.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91060568; PubMed=2246248;			
RA	Jenkins R.N., Osborne-Lawrence S.L., Sinclair A.K., Eddy R.L. Jr.,			
RT	Byers M.G., Snows T.B., Duby A.D.,			
RT	"Structure and chromosomal location of the human gene encoding			
RT	cartilage matrix protein."			

RL J. Biol. Chem. 265:19624-19631(1990).
 CC -1- FUNCTION: CARTILAGE MATRIX PROTEIN IS A MAJOR COMPONENT OF THE
 CC EXTRACELLULAR MATRIX OF NONARTICULAR CARTILAGE. IT BINDS TO
 CC COLLAGEN.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 WFMA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M55682; AAB38702.1; -.
 DR EMBL: M55675; AAB38702.1; JOINED.
 DR EMBL: M55676; AAB38702.1; JOINED.
 DR EMBL: M55677; AAB38702.1; JOINED.
 DR EMBL: M55679; AAB38702.1; JOINED.
 DR EMBL: M55680; AAB38702.1; JOINED.
 DR EMBL: M55681; AAB38702.1; JOINED.
 DR EMBL: M55683; AAB38702.1; JOINED.
 DR PIR: A37979; A37979.
 DR HSSP: P05099; IAOS.
 DR Gene: HGNC:6907; MATN1.
 DR MIM: 115437; -.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002035; WVF_A.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF00092; Wvf; 2.
 DR PRINTS: PRO0453; WVFADOMAIN.
 DR SMART: SM00181; EGF; 1.
 DR SMART: SM00327; WVF; 2.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS0186; EGF_2; 1.
 DR PROSITE: PS50234; WVF; 2.
 KW EGF-like domain; signal; Glycoprotein; Cartilage; Repeat; Coiled coil.
 FT SIGNAL 1 222
 FT CHAIN 23 496
 FT DOMAIN 223 222
 FT DOMAIN 223 222
 FT DOMAIN 223 222
 FT DOMAIN 223 222
 FT DOMAIN 223 222
 FT DOMAIN 223 222
 FT CARBOHYD 76 76
 FT DISULFID 35 221
 FT DISULFID 227 238
 FT DISULFID 234 247
 FT DISULFID 249 262
 FT DISULFID 265 452
 FT DISULFID 265 452
 SQ SEQUENCE 496 AA; 53700 MW; 2D880A8114C7940F CRC64;
 Query Match 20.0%; Score 205; DB 1; Length 496;
 Best Local Similarity 33.5%; Pred. No. 4,3e-11;
 Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

AC Q91145;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XII) chain (Fragment).
 OS Notophthalmus viridescens (Eastern newt) (Triturus viridescens).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;
 OC Notophthalmus.
 OK NCBI_Taxid=8316;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95246925; PubMed=7729585;
 RA Wei Y., Yang E.V., Klatz K.P., Tassava R.A.;
 RT "Monoclonal antibody M72 identifies the urodele alpha 1 chain of type
 RT XII collagen, a developmentally regulated extracellular matrix
 RT protein in regenerating newt limbs.";
 RL Dev. Biol. 168:503-513(1995).
 CC -1- FUNCTION: TYPE XII COLLAGEN INTERACTS WITH TYPE I COLLAGEN-
 CC CONTAINING FIBRILS, THE COIL DOMAIN COULD BE ASSOCIATED WITH THE
 CC SURFACE OF THE FIBRILS, AND THE COIL AND NC3 DOMAINS MAY BE
 CC LOCALIZED IN THE PERIFIBRILLAR MATRIX (BY SIMILARITY). COULD PLAY
 CC A DEVELOPMENTAL ROLE IN REGENERATION.
 CC -1- SUBUNIT: TRIMER OF IDENTICAL CHAINS EACH CONTAINING 190 kDa OF
 CC NONTRIPLE-HELICAL SEQUENCES (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 3 DAYS AFTER AMPUTATION
 CC IN CELLS OF THE BASAL LAYER OF THE WOUND EPITHELIUM. AT DAY 10,
 CC EXPRESSION IS FOUND IN BOTH THE BASAL WOUND EPITHELIAL CELLS AND
 CC THE DISTAL MESenchyme CELLS. AT MID-BUD AND LATE-BUD BLASTOMA
 CC STAGES, WOUND EPITHELIUM EXPRESSION HAS DECREASED, WHEREAS THE
 CC MESenchyme REMAINS STRONGLY ACTIVE IN TRANSCRIPTION AND SHOWED A
 CC TENDENCY TOWARD DISTAL REGIONALIZATION. CONDENSING CARTILAGE SHOWS
 CC NO SIGNAL. FINALLY, AT THE LATE DIGIT STAGE, EXPRESSION BECOMES
 CC LARGELY RESTRICTED TO THE PERICHONDRUM.
 CC -1- PTM: THE TRIPLE-HELICAL TAIL IS STABILIZED BY DISULFIDE BONDS AT
 CC EACH END (BY SIMILARITY).
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 WFMA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U19494; AAB0217.1; -.
 DR HSSP: P02751; IENA.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR002035; WVF_A.
 DR Pfam: PF00041; fn3; 7.
 DR Pfam: PF00092; Wvf; 2.
 DR Pfam: PF00092; Wvf; 2.
 DR SMART: SM00060; FN3; 5.
 DR SMART: SM00327; WVF; 1.
 DR PROSITE: PS50234; WVF; 2.
 KW Extracellular matrix; Connective tissue; Repeat; Cell adhesion;
 KW Collagen; Glycoprotein.
 FT NON_TER 1 1
 FT DOMAIN 1 49
 FT DOMAIN 63 154
 FT DOMAIN 155 245
 FT DOMAIN 246 338
 FT DOMAIN 339 432
 FT DOMAIN 433 519
 FT DOMAIN 520 612
 FT DOMAIN 633 805
 FT DOMAIN 818 907
 FT DOMAIN 908 >929
 WVF 1.
 FIBRONECTIN TYPE-III 1.
 FIBRONECTIN TYPE-III 2.
 FIBRONECTIN TYPE-III 3.
 FIBRONECTIN TYPE-III 4.
 FIBRONECTIN TYPE-III 5.
 FIBRONECTIN TYPE-III 6.
 WVF 2.
 FIBRONECTIN TYPE-III 7.
 FIBRONECTIN TYPE-III 8.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 20.3291 Seconds
(without alignments)
2006.842 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337
Perfect score: 1025
Sequence: 1 GICARVDASFPQGSIAFTA.....FLIREIRIASDPDERFFPNV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	473	46.1	1171	13 Q42094	Q42094 gallus gall
2	336	32.8	191	6 Q29124	Q29124 sus scrofa
3	207.5	20.2	1472	13 Q90207	Q90207 gallus gall
4	205.5	20.0	1207	4 Q9B0U7	Q9B0U7 homo sapien
5	204.5	20.0	1140	4 Q9P218	Q9P218 homo sapien
6	199	19.4	3767	5 Q9UAI3	Q9UAI3 caenorhabdi
7	196	19.1	956	11 Q8R542	Q8R542 mus musculu
8	191	18.6	937	4 Q96FT5	Q96FT5 homo sapien
9	191	18.6	956	11 Q99K64	Q99K64 mus musculu
10	188.5	18.4	2104	5 Q21281	Q21281 caenorhabdi
11	188.5	18.4	2104	5 Q964N4	Q964N4 caenorhabdi
12	182.5	17.8	526	4 Q9Y3N2	Q9Y3N2 homo sapien
13	182.5	17.8	954	4 Q8XV8	Q8XV8 homo sapien
14	182.5	17.8	957	4 Q96P44	Q96P44 homo sapien
15	182	17.8	387	11 Q60863	Q60863 mus musculu
16	181.5	17.7	957	4 Q9H0V3	Q9H0V3 homo sapien

17	181	17.7	549	6 Q02808	Q02808 bos taurus
18	177.5	17.3	360	4 Q9UGC3	Q9UGC3 homo sapien
19	177	17.3	664	4 Q19013	Q19013 erinaceus e
20	175	17.1	724	5 Q04588	Q04588 eimeria max
21	172.5	16.8	386	11 Q91YC3	Q91YC3 otomyx ang
22	170.5	16.6	421	11 Q9UK16	Q9UK16 mus musculu
23	170.5	16.6	843	4 Q05707	Q05707 homo sapien
24	169	16.6	537	4 Q96AA0	Q96AA0 homo sapien
25	169	16.5	1169	11 Q90XH4	Q90XH4 mus musculu
26	168.5	16.4	200	11 Q88494	Q88494 mus musculu
27	168	16.4	1703	11 Q92019	Q92019 mus musculu
28	167	16.3	517	4 Q43853	Q43853 homo sapien
29	165.5	16.1	382	11 Q91Y33	Q91Y33 jaculus jac
30	165.5	16.0	385	11 Q91YF6	Q91YF6 micromys mi
31	164.5	16.0	410	11 Q92114	Q92114 allactaga e
32	163.5	16.0	380	11 Q91YH0	Q91YH0 calomyscus
33	163	15.9	1253	11 Q97566	Q97566 canis fami
34	162.5	15.9	411	6 Q9GL88	Q9GL88 inia geoffr
35	161.5	15.8	417	11 Q920Q9	Q920Q9 dryomys nlt
36	161	15.7	755	4 Q00261	Q00261 homo sapien
37	159.5	15.6	385	11 Q91YH1	Q91YH1 cricetus
38	159.5	15.6	385	11 Q91YH1	Q91YH1 lophuromys
39	159.5	15.6	385	11 Q91YH1	Q91YH1 photopus ro
40	159.5	15.6	385	11 Q91YH1	Q91YH1 tatera kemp
41	159	15.5	393	6 Q9GL87	Q9GL87 mesopiodon
42	158.5	15.5	406	6 Q77672	Q77672 scalopus aq
43	158.5	15.5	421	11 Q9EPN7	Q9EPN7 heliophobiu
44	158	15.4	412	6 Q9MZE6	Q9MZE6 ziphius cav
45	158	15.4	1160	11 Q9R200	Q9R200 mus musculu

ALIGNMENTS

RESULT 1	
Q42094	PRELIMINARY; PRT; 1171 AA.
AC Q42094;	
DT 01-JAN-1998 (TREMBLrel. 05, Created)	
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)	
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)	
DE ALPHAI Integrin.	
OS Gallus gallus (Chicken).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC Gallus.	
OX NCBI_TaxID=9031;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=GITZARD;	
RX MEDLINE=97476270; PubMed=9334246;	
RA Obata H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,	
RA Sobue K.;	
RT "Smooth muscle cell phenotype-dependent transcriptional regulation of	
RT the alpha integrin gene.";	
RL J. Biol. Chem. 272:26643-26651 (1997).	
EMBL AB000470; BAA23160.1; -	
EMBL AB000471; BAA23161.1; -	
HSSP; P17301; IAOX.	
InterPro; IPR000413; Integrin_alpha.	
InterPro; IPR002035; VWF_A.	
Pfam; PF01839; FG-GAP; 5	
Pfam; PF00357; Integrin_A; 1.	
Pfam; PF00092; vwa; 1.	
PRINTS; PR00453; VWFADOMAIN.	
SMART; SM00191; Int_alpha; 4.	
SMART; SM00327; VWA; 1.	
PROSITE; PS00242; INTEGRIN_ALPHA; 1.	
PROSITE; PS50234; VWA; 1.	
Integrin.	
SEQUENCE 1171 AA; 130228 MW; B505A4D65F09736E CRC64;	
Query Match	46.1%; Score 473; DB 13; Length 1171;

Best Local Similarity 45.5%; Pred: No. 5.3e-35;
Matches 90; Conservative 37; Mismatches 71; Indels 0; Gaps 0;

Qy 1 GICARVADSPQPGSLAPTAORCEYMDVIVLDGNSNLYPMSVEQFLRLRYOKLIDP 60
Db 135 GVCNSVNSTEYTKAAPSVQCEKTDIVIVLDGNSNLYPMSVTAFLNLSNMIDIGP 194
Qy 61 EQLQVGLVQGESPEVHWSLGDPRTEEVEVRAAKULSREGRGETYTAQAIWACTEGFSQ 120
Db 195 QQVQVGVVQVQGVVHVEFYNTSYSTEEVMDAALIRQGRQIMTALGIDTAEEAFTE 254
Qy 121 SHGRPRPAALLVVTTDGSBHDEELPAALKACEAGRVTRGIVAGLYLRQDPSSFL 180
Db 255 AHGRARVQCKWMIYVTDGESHVDYRLQEVIDKCEDENIORFAIILGYSYSGNLSTKEFV 314
Qy 181 REIRTIASDPEREFNFNV 198
Db 315 EELKSIASKPTEGFNFNV 332

RESULT 2	ID	Q29124	PRELIMINARY;	PTT;	191 AA.
AC	Q29124;				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)			
DE	VLA-2 (Fragment).				
OS	Sus scrofa (Pig).				
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.				
OX	NCBI_TaxId=9823;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=ENDOTHELIAL CELLS;				
RX	MEDLINE=95036279; PubMed=7949129;				
RA	Bahou W.F., Potter C.L., Mirza H.;				
RT	"The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific				
RL	recognition sequence for endothelial cell attachment and spreading				
BL	molecular and functional characterization.";				
DR	Blood 84:3734-3741(1994).				
DR	EMBL; Z12137; CAAT8125.1; -				
DR	HSSP; P17301; 1A0X.				
DR	InterPro; IPR002035; VWF_A.				
DR	Pfam; PF00092; vwa; 1.				
DR	PRINTS; PR00453; VWFADOMAIN.				
DR	SMART; SM00327; VWA; 1.				
DR	PROSITE; PSS0234; VWFA; 1.				
FT	NON_TER	1			
FT	NON_TER	1			
FT	NON_TER	191			
FT	NON_TER	191			
QO	SEQUENCE	191 AA;	20896 MW;	2E228B472EC699D8	CRC64;

DT	01-DEC-2001	(TREMBLrel. 19, Created)
DT	01-DEC-2001	(TREMBLrel. 19, Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE	Collagen type XX alpha 1 precursor.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Atherinomorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_TaxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21303548; PubMed=11274142;	
RA	Koch M., Foley J.E., Hahn R., Zhou P., Burgess R.E., Gerecke D.R.,	
RA	Gordon M.K.	
RT	"alpha 1(XI) Collagen, a New Member of the Collagen Subfamily, Fibril-	
RT	associated Collagens with Interrupted Triple Helices."	
RL	J. Biol. Chem. 276:23120-23126(2001).	
DR	EMBL; AF312825; AAKS8847.1; -	
DR	InterPro; IPRO00087; Collagen.	
DR	InterPro; IPRO03961; FN III.	
DR	InterPro; IPRO03129; TSPN.	
DR	InterPro; IPRO02035; VWF_A.	
DR	Pfam; PF01391; Collagen; 4.	
DR	Pfam; PF00041; fn3; 6.	
DR	Pfam; PF02210; TSPW; 1.	
DR	Pfam; PF00092; Vwa; 1.	
DR	ProDom; PD000007; Collagen; 1.	
DR	ProSITE; PSS0234; VMPA; 1.	
KW	Signal.	
FT	SIGNAL 1 28	POTENTIAL.
SQ	SEQUENCE 1472 AA; 156903 MW; 5361611579C56EFD CRC64;	

RESULT 3
Q90ZA0
ID Q90ZA0 PRELIMINARY; PRT; 1472 AA
AC Q90ZA0;

RESULT 4		
09BOUT7	PRELIMINARY;	PRT, 1207 AA.
ID	09BOUT7.	
AC	01-JUN-2001 (TREMBlrel. 17, Created)	
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)	
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)	
DE	BA261N11.4.1 (K1A1A1510, isoform 1).	
GN	BA261N11.4.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxID=9606;	
RV	[1]	
RP	SEQUENCE FROM N.A.	
RA	Hall R.;	
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AL121827; CAC36117.1; -	
DR	HSSP; P17301; IAOX.	
DR	InterPro; IPR0000087; Collagen.	

DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen_1.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; vwa; 1.
DR PROSITE; PS50234; VWF_A; 1.
SQ SEQUENCE 1207 AA; 128085 MW; 368158C26C44F1C1 CRC64;

Query Match 20.0%; Score 205.5; DB 4; Length 1207;
Best Local Similarity 29.5%; Pred. No. 3.5e-10;
Matches 59; Conservative 37; Mismatches 71; Indels 33; Gaps 7;

Qy 11 QPGSLAFTAQ-----RC--PTYMDVIVLDGNSI--YPMSEVQTFLR 50
Db 151 EPQVAFTSQDPRTPGSEWRETGPQFRLPVPADWFLVDGWSISGSHFQVQKDFLA 210
Qy 51 RLVGKLFIDPEQIQGLVQYGSFVHWSLQDFRTKEVYRAAKLSRREGRTTAQAI 110
Db 211 SVIAPFEIGPKVQVGLTQYSGDAQTEWDLNSLSTKEQVLAVRRL-RYKGGNTFTGLAL 269
Qy 111 MVACTGEFSQSHGSPPEARLLVVTDESHDGEELPALKACEAGRYTRYGIAVLGHYLL 170
Db 270 THVLGQNIQPAAGLRFPEAKKVIILTDGKSD--DVHTAARVLKDLGVNFAVGV----- 322
Qy 171 RRQRDPSSFLEIRITIASDP 190
Db 323 -KNADEA---ELRLLASPP 337

RESULT 5

Q9P218 PRELIMINARY; PRT; 1140 AA.
AC Q9P218;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE KIAA1510 protein (Fragment).
GN KIAA1510.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.,
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro".
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040943; BAA96034.1; -
DR HSBP; P17301; IAOX.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen_2.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00060; FN3; 5.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00327; vwa; 1.
DR PROSITE; PS50234; VWF_A; 1.
FT NON_TER 1 1
SQ SEQUENCE 1140 AA; 120478 MW; 9EABD7D24FD67D4D CRC64;

Query Match 20.0%; Score 204.5; DB 4; Length 1140;
Best Local Similarity 32.4%; Pred. No. 4e-10;
Matches 56; Conservative 34; Mismatches 66; Indels 17; Gaps 6;

Qy 22 RC--PTYMDVIVLDGNSI--YPMSEVQTFRLRLVGKLFIDPEQIQGLVQYGSFVH 77
Db 21 RCLPVPADWFLVDGWSISGSHFQVQKDFLASVIAPFEIGPKVQVGLTQYSGDAQTE 80
Qy 78 WSLGDFRTKEVYRAAKLSRREGRTTAQAIWVACTGEFSQSHGSPPEARLLVVTVD 137
Db 81 WDLNSLSTKEQVLAVRRL-RYKGGNTFTGLALTHVLGQNIQPAAGLRFPEAKKVIILTD 139
Qy 138 GESHGDELPAALKACEAGRYTRYGIAVLGHYLRQRPSSFLREIRITIASDP 190
Db 140 GKSD--DVHTAARVLKDLGVNFAVGV-----KNADEA---ELRLLASPP 180

RESULT 6

Q9UAL3 PRELIMINARY; PRT; 3767 AA.
AC Q9UAL3; Q21340;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Transmembrane cell adhesion receptor MUA-3 precursor (K08E5.3 protein)
DE (Fragment).
GN MUA-3 OR K08E5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitida; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2.
RA Lu Z., Vogel B., Hedgecock E.;
RT "mu-3 mRNA Splicing Pattern Revealed";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kershaw J.K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [4]
RP SEQUENCE FROM N.A.
RA Berkis M., Smith A., Lloyd C.R.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF119060; AAD29428.1; -
DR EMBL; Z30974; CAAB3826.2; -
DR EMBL; Z30423; CAAB3826.2; JOINED.
DR EMBL; Z30423; CAC42345.1; -
DR EMBL; Z30974; CAC42345.1; JOINED.
DR HSBP; P01130; IIDL.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR002172; LDL_recept_A.
DR InterPro; IPR000082; SEA_domain.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 33.
DR Pfam; PF00057; Idl_recept_a; 3.
DR Pfam; PF01390; SEA; 2.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PRINTS; PR00453; VWFADOMAIN.

```

DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_Like; 45.
DR SMART; SM00192; LDLA; 4.
DR SMART; SM00200; SEA; 2.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_32.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS00068; LDLA_2; 2.
DR PROSITE; PS00024; SEA; 4.
DR PROSITE; PS00234; VMA; 2.
DR Calcium-binding; EGF-like domain; Glycoprotein; Receptor; Repeat;
KW Signal.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN 25 >3767 3. TRANSMEMBRANE CELL ADHESION RECEPTOR MUA-
FT NON_TER 3767 3767
SQ SEQUENCE 3767 AA; 417284 MW; 8DA3AE5EA50AE8BE CRC64;

Query Match
Best Local Similarity 19.4%; Score 199; DB 5; Length 3767;
Matches 59; Conservative 33; Mismatches 83; Indels 28; Gaps 8;

QY 6 VDASFO---PQSLAFTAQCPTV--MDVYIVLDGNSIYPM--SEVQTFELRLVGLKFI 58
DB 1204 VDVSSNANLPGRVCTVQTTCPKQKTDLVFLIDSGSIGSYVFKNEVLFVAFELFEI 1263
QY 59 DEBOIQVGLVOYGESPVHEWSLGDPRTEKEVYRAANKNSRRREGRETKTAQIMVACTEGF 118
DB 1264 GRSKTRVGLIOYSDQIRHFEFDLDQYGDRLSLKLGISETOYLTGL--TRTAAIIOHWQEGF 1322
QY 119 SQSHGRP---EAARLLVVVTGDGSHDGEELPAALACAGR--VTRYGIVLGHYLR 172
DB 1323 SERGARPPQSDARVAIILIDRSODNVTGPA-----DSARKLSINTPAIGTDHVL-- 1375
QY 173 QRPDSFLREIRTIASDPDERFF 195
DB 1376 -----ASELESIAQSPNRWFY 1391

RESULT 7
ID Q8R542 PRELIMINARY; PRT; 956 AA.
AC Q8R542;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE Matrilin-2.
GN MATN2.
OS Mus musculus (Mouse).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21841526; PubMed=11852232;
RA Matas L., Korpos E., Deak F., Liu Z., Beier D.R., Aszodi A., Kiss I.;
RT "Comparative analysis of the mouse and human genes (Matn2 and MATN2)
RT for matrilin-2, a filament-forming protein widely distributed in
RT extracellular matrices."
RT Matrix Biol. 21:163-174(2002).
DR EMBL; AF358844; AAM1539.1; -.
DR EMBL; AF358831; AAM1539.1; JOINED.
DR EMBL; AF358832; AAM1539.1; JOINED.
DR EMBL; AF358833; AAM1539.1; JOINED.
DR EMBL; AF358834; AAM1539.1; JOINED.
DR EMBL; AF358835; AAM1539.1; JOINED.
DR EMBL; AF358836; AAM1539.1; JOINED.
DR EMBL; AF358837; AAM1539.1; JOINED.
DR EMBL; AF358838; AAM1539.1; JOINED.
DR EMBL; AF358839; AAM1539.1; JOINED.

```

```

DR EMBL; AF358840; AAM1539.1; JOINED.
DR EMBL; AF358841; AAM1539.1; JOINED.
DR EMBL; AF358842; AAM1539.1; JOINED.
DR EMBL; AF358843; AAM1539.1; JOINED.
SQ SEQUENCE 956 AA; 106748 MW; 3E4A608FF92BDE55 CRC64;

Query Match
Best Local Similarity 19.1%; Score 196; DB 11; Length 956;
Matches 64; Conservative 33; Mismatches 83; Indels 32; Gaps 8;

QY 6 VDASFOQSGSLAFTAQCPTV-----MDVYIVLDGNSI--YPMSEVQTF 49
DB 21 VDGRRERPPARPPSRGRHWRWYPTALLSSCENKRAIDLFIIDSSRSVNTDYAKVKEFI 80
QY 50 RLTVKLEIDPEQIOVGLVOYGESPVHEWSLGDPRTEKEVYRAANKNSRRREGRETKTAQ 109
DB 81 LDLIQFLIDIPDVRVGLVIGSTVKNFSLKTFKRSSEVERAYKRW-RHLSTGTMTGLA 139
QY 110 IMVACTEGFSQSHGRP---EAARLLVVVTGDGSHDGEELPAALACAGRVTTRYGI 166
DB 140 IOYALNINFSAEAGRPARENVPRILMIVTDGRQDVAEYVA-KARNTG-ILIFALGV- 196
QY 167 GHYLRQRPDSFLREIRTIASDPDERFFNV 198
DB 197 -----GQVD-----INTLKAIGSEBPKDHVFLV 219

RESULT 8
ID Q96FT5 PRELIMINARY; PRT; 937 AA.
AC Q96FT5;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Matrilin 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010444; AAH10444.1; -.
DR InterPro; IPR000152; ASX_hydroxyl.
DR InterPro; IPR000561; EGF_Like.
DR InterPro; IPR01881; EGF_CA.
DR InterPro; IPR02035; VWF_A.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00092; VWA; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_9.
DR PROSITE; PS01186; EGF_2; UNKNOWN_9.
DR PROSITE; PS0234; VMA; 2.
SQ SEQUENCE 937 AA; 104775 MW; 7C76C5CE70F06E40 CRC64;

Query Match
Best Local Similarity 18.6%; Score 191; DB 4; Length 937;
Matches 58; Conservative 33; Mismatches 67; Indels 18; Gaps 7;

QY 28 DVVIVLDGNSI--YPMSEVQTFRLRVGLKFIIDPEQIOVGLVOYGESPVHEWSLGDPR 85
DB 57 DLVFIIDSSRSVNHDAKVEFIVDLIQFLIDIPDVRVGLVIGSTVKNFSLKTFK 116
QY 86 KEVYRAANKNSRRREGRETKTAQIMVACTEGFSQSHGRP---EAARLLVVVTGDGSHD 142
DB 117 KSEVERAYKRW-RHLSTGTMTGLAIOYALNINFSAEAGARLRENVPRIVIMIVTDGRQD 175
QY 143 GEEIPALACAGRVTTRYGIAGVYLRQRPDSFLREIRTIASDPDERFFNV 198
DB 176 SVAEYVA-KARNTG-ILIFALGV-----GQVD-----FNTLKSIGSEBHDHVLV 219

```

```

RESULT 9
O99K64 ID O99K64 PRELIMINARY; PRT; 956 AA.
AC O99K64;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Matrilin 2.
GN MATN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005429; AA05429.1; -.
DR HSSP; P05099; IA05.
DR MGD; MGI:109613; Matn2.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 10.
DR Pfam; PF00092; vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF_CA; 10.
DR SMART; SM00179; EGF_CA; 10.
DR SMART; SM00327; vwa; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 7.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50234; VWFA; 2.
KW EGF-like domain; Glycoprotein; Hydroxylation.
SQ SEQUENCE 956 AA; 106706 MM; 373BFC774370BFEE CRC64;

Query Match 18.6%; Score 191; DB 11; Length 956;
Best Local Similarity 29.7%; Pred. No. 5,7e-09;
Matches 63; Conservative 34; Mismatches 83; Indels 32; Gaps 8;

OY 6 VDASQPGGSLAPTRQRCPT;-----MDVYIVLDGNSI--YPMSEVQFL 49
DB 21 VDGRRPQARFPPSRGRHVRMPQOTALLSSCENKRAIDIVFIIDSSRSVTHDYAKVKEFI 80
OY 50 RLVLKLTIDPEQIOVGLVQYGESPVHEMSLDGDFRTKEEVRAAKNLSRREGRETKTQDA 109
DB 81 LDILQFLDIGPVTVTGVLQIGSTYVKNFSLTKTFKSKSEVERAVKRM-RHLSTGTMGLA 139
OY 110 IMVACTEGFSQSHGRP--EAARLLVVTDDGSHDGEELPAALKACEAGRTVRYGIAVL 166
DB 140 IQVALNINAFSEAEGRAPLRENVPRITIMIVTDGRPDQSVAEVA--KARNTG-ILIPALGV- 196
OY 167 GHYLRQRPDPSFLREIRTIASDPPERFFENV 198
DB 197 -----GVVD---LNTLKAIGSEPHKDHVFLV 219

RESULT 10
O21281 ID O21281 PRELIMINARY; PRT; 2104 AA.
AC O21281;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 230.1 kDa protein.
GN K07D8.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;

```

```

RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Favello T.;
RT "The sequence of C. elegans cosmid K07D8.";
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; L16679; AAA28092.5; -.
DR HSSP; P02468; I1LE.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR000082; SEA_domain.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00008; EGF; 20.
DR Pfam; PF00390; SEA; 4.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00181; EGF; 35.
DR SMART; SM00179; EGF_CA; 24.
DR SMART; SM00200; SEA; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS50024; SEA; 4.
DR PROSITE; PS50234; VWFA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hypothetical protein;
KW Repeat.
SQ SEQUENCE 2104 AA; 230051 MM; C98CD462CC43F8B CRC64;

Query Match 18.4%; Score 188.5; DB 5; Length 2104;
Best Local Similarity 28.5%; Pred. No. 2.8e-08;
Matches 57; Conservative 33; Mismatches 85; Indels 25; Gaps 8;

OY 6 VDASQPGGSLAPTRQRCPT;MDVYIVLDGNSIYPM--SEVQFLRLVLKGLFI 58
DB 411 VDSSNNMLPGRGCTCTSTACPAQPTDLVFLIDSGSIGSVFQTEVLRFLAEFTLELFDI 470
OY 59 DPEQIOVGLVQYGESPVHEMSLDGDFRTKEEVRAAKNLSRREGRETKTAQIMVACTEGF 118
DB 471 APQTRRSVVOYSQIIRHEFELDYNVSRKSIQNAIRINIEYTG-TRTGAALIEVNAEAF 529
OY 119 SQSHGRP--EAARLLVVTDDGSHDGEELPAALKACEAGR--VTRYGIAVLGHYLRQR 174
DB 530 SERRGARPVGVSVAVIYTDGRSQDNVTRPSD-----NARQDILQLFAVGTNHYLD--- 582
OY 175 DPSSFLREIRTIASDPPERF 194
DB 583 -----AELEISGSKDRTF 596

RESULT 11
O964N4 ID O964N4 PRELIMINARY; PRT; 2104 AA.
AC O964N4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Transmembrane matrix receptor MUP-4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

```

OC Rhabdittidae; Peloderinae; Caenorhabdittis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hong L., Elbl T., Franzini-Armstrong C., Ward J., Rybicka K.K.,
 RA Gatewood B.K., Bucher E.A.;
 RT "MUP-4 is a novel matrix receptor with essential functions in
 RT epithelial cell adhesion at hemidesmosomes and transmission of muscle
 RT forces in Caenorhabdittis elegans";
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF289202; AAK69172.1;
 DR EMBL; AF289202; AAK69172.1;
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF Ca.
 DR InterPro; IPR000082; SEA domain.
 DR Pfam; PF00008; EGF_16.
 DR Pfam; PF00092; VWA; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_13.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_3.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_3.
 DR PROSITE; PS01187; EGF_CA; UNKNOWN_2.
 DR PROSITE; PS50024; SEA; 2.
 DR PROSITE; PS50234; VMPA; 1.
 KW Receptor.
 SQ SEQUENCE 2104 AA; 230035 MW; 31174C0CCDB61CC2 CRC64;
 Query Match 18.4%; Score 188.5; DB 5; Length 2104;
 Best Local Similarity 28.5%; Pred. No. 2.8e-08;
 Matches 57; Conservative 33; Mismatches 85; Indels 25; Gaps 8;
 QY 6 VASPFQ---PGSLAPTAQRCPY--MDVIVIVLDGNSIYP--SEVQTFRLRVGLKFI 58
 DB 411 VVSSNANLPGRVCTLTACPAQPTDVLFLIDGSSISYFQTEVLEFLAEFFELFPI 470
 QY 59 DEPOIQVGLVQYGESEPHWMSLGFPTKEEVRAAKNLSRREGRETKTAQAIMVACTEGF 118
 DB 471 AQOKRVSVVQYSDQIRHFEGLDNVSGDRSLQNAINIIYGL-TRTAAIEHVANNAF 529
 QY 119 GSCHGRP--EAARLIVVTGESHGDELPALAKACEGR--VTRYGIADVGHYLROR 174
 DB 530 SERRGARVPQVSRVAIVITDGSQDNVTRPSD---NARODIQLEFAGVNTNHYLD-- 582
 QY 175 DPSSFLEIRTIASDPDERF 194
 DB 583 -----ALDEIISGSKDRTF 596
 RESULT 12
 OY3N2 PRELIMINARY; PRT; 526 AA.
 ID OY3N2;
 AC OY3N2;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE DJ708FS.1 (Putative novel collagen alpha 1 like protein)
 DE (Fragment).
 GN DJ708FS.1.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smalley C.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031782; CAB40275.1;
 DR HSSP; PI7301; IAOX.
 DR InferPro; IPR000087; Collagen.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01391; Collagen; 2.

DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50234; VMPA; 1.
 KW Collagen.
 FT NON TER 526
 SQ SEQUENCE 526 AA; 57189 MW; 44016B504F8A5E1 CRC64;
 Query Match 17.8%; Score 182.5; DB 4; Length 526;
 Best Local Similarity 30.5%; Pred. No. 1.6e-08;
 Matches 57; Conservative 36; Mismatches 75; Indels 19; Gaps 6;
 QY 14 GSLAPTAQRCPYMDVIVIVLDGNSIYP--MSEVQTFRLRVGLKFIPEQIQVGLVQY 71
 DB 25 GEVRSSCTAPT--DLVFILDGSYSGPENEIVKMLVNTKFNFDIPKTIQGVQVYS 82
 QY 72 ESPVHEWSLADPRTKEEVRAAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARL 131
 DB 83 DYPVLEIFLGSDGSEHLLTAVESILYL-GGNTKTKGAIQFALDYLFKAS--SRFLTKI 138
 QY 132 LVVVTGSHGDELPALAKACEGRVTRYGIADVGHYLRORDBSSFLEIRTIASDPD 191
 DB 139 AVVLTGKSD--DVKDAQAARDSKITLFAIGV-----GSEIDAEIRAIANKPS 187
 QY 192 ERPFENV 198
 DB 188 STVPEYV 194
 RESULT 13
 OY3N2 PRELIMINARY; PRT; 954 AA.
 ID OY3N2;
 AC OY3N2;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Alpha 1 chain-like collagen COL1A1 precursor.
 GN COL1A1.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chou M.-Y.;
 RT "Cloning and identification of a novel human alpha 1 chain-like
 RT collagen (COL1A1) gene, a new member of the FACIT family";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF330693; AAL50033.1;
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR003129; TSPN.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF02210; TSPN; 1.
 DR Pfam; PF01391; Collagen; 6.
 DR Pfam; PF00092; VWA; 1.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR ProDom; PD000007; Collagen; 4.
 DR SMART; SM00210; TSPN; 1.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS50234; VMPA; 1.
 DR Signal; Collagen.
 FT SIGNAL 1
 SQ SEQUENCE 954 AA; 99068 MW; C3514675DADCF23 CRC64;
 Query Match 17.8%; Score 182.5; DB 4; Length 954;
 Best Local Similarity 30.5%; Pred. No. 3.5e-08;
 Matches 57; Conservative 36; Mismatches 75; Indels 19; Gaps 6;
 QY 14 GSLAPTAQRCPYMDVIVIVLDGNSIYP--MSEVQTFRLRVGLKFIPEQIQVGLVQY 71
 DB 25 GEVRSSCTAPT--DLVFILDGSYSGPENEIVKMLVNTKFNFDIGPKTIQGVQVYS 82

QY 72 ESPVHEWISLGDFTKEEVVRAKNLSRRRGRETQTAQAIMVACTEGFSQSGHGREPAARL 131
DB 83 DYPVLEIPFGSYDSEHILTAIVESILYL-GGNTKTKGKAQFALDYLFAKS---SRFLTKI 138
QY 132 LVVVTDDGSHDGEELPALKACEAGRTVRYGIAVLGHYLRORDDSSFLREIRTIASDPD 191
DB 139 AAVLTLDGKSQD--DVKDAQAARDSKITLFAIGV-----GSETEDAEIRAIANKPS 187
QY 192 ERFFENV 198
DB 188 STYVEYV 194

RESULT 14

Q96P44 PRELIMINARY; PRT; 957 AA.
AC Q96P44; 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Collagen XXI (Alpha 1 type XXI collagen precursor).
GN COL21A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RA Fitzgerald J., Bateman J.F.;
RT "A new FACIT of the collagen family: COL21A1.";
RL FEBS Lett. 0:0-0(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21853310; PubMed=11863369;
RA Chou M.Y., Li H.C.;
RT "Genomic Organization and Characterization of the Human Type XXI
RT Collagen (COL21A1) Gene.";
RN Genomics 79:395-401(2002).
DR EMBL; AF414088; AAL02227.1; -
DR EMBL; AF438327; AAL86699.1; -
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00092; vwa; 1.
DR PDBOM; PDB00007; Collagen; 4.
DR PROSITE; PS50234; VWF_A; 1.
DR Collagen; Signal.
KW SIGNAL.
FT SIGNAL 1 22
SQ SEQUENCE 957 AA; 99369 MW; 4CSODFE6656A675 CRC64; POTENTIAL.

Query Match 17.8%; Score 182.5; DB 4; Length 957;
Best Local Similarity 30.5%; Pred. No. 3.5e-08;
Matches 57; Conservative 36; Mismatches 75; Indels 19; Gaps 6;

QY 14 GSIAPTACRPTVMVIVLIDGNSIYP--WSEVQTFRLRVGKFLFIDEPQIQVGLVQYG 71
DB 25 GEVRSSCRAPF--DLVFLIDGSYSVGPENFEIVKKMLVNTKXNPDIGKPTQVAVOYS 82
QY 72 ESPVHEWISLGDFTKEEVVRAKNLSRRRGRETQTAQAIMVACTEGFSQSGHGREPAARL 131
DB 83 DYPVLEIPFGSYDSEHILTAIVESILYL-GGNTKTKGKAQFALDYLFAKS---SRFLTKI 138
QY 132 LVVVTDDGSHDGEELPALKACEAGRTVRYGIAVLGHYLRORDDSSFLREIRTIASDPD 191
DB 139 AAVLTLDGKSQD--DVKDAQAARDSKITLFAIGV-----GSETEDAEIRAIANKPS 187
QY 192 ERFFENV 198
DB 188 STYVEYV 194

RESULT 15
Q60863 PRELIMINARY; PRT; 387 AA.
ID Q60863
AC Q60863;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Von Willebrand factor (Fragment).
GN VWF OR VWF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=93305893; PubMed=8318738;
RA Barrow L.L., Simin K., Mohlke K.L., Nichols W.C., Ginsburg D.,
RA Weisler M.H.;
RT "Conserved linkage of neurotrophin-3 and von Willebrand factor on
RT mouse chromosome 6.";
RL Mamm. Genome 4:343-345(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=94250904; PubMed=8193357;
RA Nichols W.C., Cooney K.A., Mohlke K.L., Baillet J.D., Yang A.,
RA Bruck M.E., Redington M., Novak E.K., Swank R.T., Ginsburg D.;
RT "von Willebrand disease in the RIIIS/J mouse is caused by a defect
RT in Blood 1995 Sep 15;86(6):2461.";
RL Blood 83:3225-3231(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Nichols W.C., Mohlke K.L., Yang A., Bruck M.E., Ginsburg D.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U27810; AAA82929.1; -
DR HSSP; P04275; IAVO.
DR MGP; MGI:98941; VWF.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 2.
DR SMART; SMO0327; VWA; 2.
DR PROSITE; PS50234; VWF_A; 2.
FT NON_TER 1 1
FT NON_TER 387 387
SQ SEQUENCE 387 AA; 43166 MW; DTFBACS2465B950 CRC64;

Query Match 17.8%; Score 182; DB 11; Length 387;
Best Local Similarity 26.8%; Pred. No. 1.2e-08;
Matches 52; Conservative 43; Mismatches 89; Indels 10; Gaps 5;

QY 4 ARVDASFGPQSLAPTAQRCPTVMVIVLIDGNSI--YPMSEVQTFRLRVGKFLFIDEP 61
DB 177 AHVTVSPGIACTLBPGRKSMVLDVFLBGSDEVGANFNKSEFEVEIQRMDVSPD 236
QY 62 QIQVGLVQYGSYPVHEWISLGDFTKEEVVRAKNLSRRRGRETQTAQAIMVACTEGFSQ 121
DB 237 ATRISVLQSYTVTWEVAFNGAQSKEVLRHVRIIRYOGGRTVTGALQYLSHSPSPS 296
QY 122 HGRPEEARLIVVVTDDGSHDGEELPALKACEAGRTVRYGIAVLGHYLRORDDSS-- 178
DB 297 QGDVEANLVVVTGNTGNASDEIRLPEDIQVVGIG---VGPHANNQELERISRTIPI 352
QY 179 FLREIRTIASD-PD 191
DB 353 FIRDFETLPREAPD 366

Search completed: July 16, 2003, 07:55:25
Job time : 22.3291 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 22 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025
Sequence: 1 GICARVDASFOGSLAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1025	100.0	1132	21	AAV32243
2	1025	100.0	1152	22	AAAB6457
3	1025	100.0	1152	22	AAAB6458
4	1025	100.0	1167	21	AAV32242
5	1025	100.0	1167	22	AAAB64584
6	912	89.0	195	23	AAU76853
7	912	89.0	195	23	AAU76862
8	797.5	77.8	303	21	AAV32282
9	547	53.4	1188	22	AAAB50087
10	547	53.4	1188	23	AAU10552

11	544	53.1	707	22	AAU19663
12	544	53.1	1034	21	ABP47883
13	544	53.1	707	23	AAAB2590
14	544	53.1	1188	22	AAU14231
15	544	53.1	1188	22	AAU14467
16	544	53.1	1188	22	AAAB30929
17	544	53.1	1188	22	AAAB50085
18	544	53.1	1188	23	AAAB50551
19	544	53.1	1189	21	AAAB25582
20	544	53.1	1189	22	ABG12949
21	492	48.0	193	23	AAU76854
22	492	48.0	193	23	AAU76855
23	467.5	45.6	1179	23	ABAB90759
24	465.5	45.4	1180	23	ABAB90788
25	452.5	44.1	214	22	AAAB50041
26	449.5	43.9	214	22	AAAB50042
27	435	42.4	1367	19	AAW70542
28	430.5	42.0	1183	20	AAV07728
29	419	40.9	1183	20	AAV07728
30	417	40.7	195	23	AAU76851
31	417	40.7	195	23	AAU76860
32	389	38.0	195	23	AAU76852
33	389	38.0	195	23	AAU76861
34	374	36.5	185	22	AAU09125
35	292	28.5	979	22	ABG29239
36	285.5	27.9	148	22	AAU19634
37	285.5	27.9	148	22	AAU19794
38	285.5	27.9	148	23	ABP47854
39	285.5	27.9	148	23	ABP48014
40	282	27.5	103	22	AAU87675
41	282	27.5	103	22	AAU19822
42	282	27.5	103	22	ABP48042
43	244	23.8	176	23	ABB78816
44	209.5	20.4	3063	23	ABAB90762
45	209.5	20.4	3118	22	AAU27790

ALIGNMENTS

RESULT 1	
AAV32243	
ID	AAV32243 standard; Protein, 1132 AA.
XX	
AC	AAV32243;
XX	
DT	15-FEB-2000 (first entry)
XX	
DE	Human integrin subunit alpha-10 splice variant.
XX	
KW	Integrin alpha-10; ISA10; human; trunna; rheumatoid arthritis;
KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis;
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;
KW	fibroblast; vaccine; marker; splice variant.
OS	Homo sapiens.
XX	
EH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..22
FT	/note= "signal peptide"
FT	23..1132
FT	/note= "mature protein"
XX	
PN	MO9951639-A1.
XX	
PD	14-OCT-1999.
XX	
PF	31-MAR-1999; 99WO-SE00544.
PR	02-APR-1998; 98SF-0001164.
PR	28-JAN-1999; 99SF-0000319.
XX	
PA	(ACT1-) ACTIVE BIOTECH AB.

Human novel extrac
Human polypeptide
Protein encoded by
Human novel protei
Human novel protei
Amino acid sequenc
Human A259. Homo
Human A259 polypep
ITGAI1 protein enc
Novel human diagno
Human integrin alp
Human integrin alp
Human Tumour Endoth
Rat Tumour Endothe
Rat alpha1 integr
Human alpha1 integ
Integrin alpha-2
Armenian hamster a
Armenian hamster a
Human integrin alp
Human integrin alp
Human integrin alp
Human integrin alp
Human integrin alp
Human integrin alp
Human novel diagno
Human novel extrac
Human novel extrac
Human polypeptide
Human polypeptide
Novel central nerv
Human novel extrac
Human polypeptide
Human polypeptide
Von Willebrand fac
Human Tumour Endoc
Human full-length

Query Match	Best Local Similarity	100.0%	Score 1025	DB 21	Length 1132
Matches 198	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
QY	1	GICARVDSFQFGSLAPTAORCPTVMYDVIVYLDGNSNTPVSEVOTFLRRVLGKFLDP	60		
DB	140	GICARVDSFQFGSLAPTAORCPTVMYDVIVYLDGNSNTPVSEVOTFLRRVLGKFLDP	199		
QY	61	EQIQVGLVQYGESPVHEWSLGDFTKKEVYRAAKULSRREGRETQAQIMVACTGESFQ	120		
DB	200	EQIQVGLVQYGESPVHEWSLGDFTKKEVYRAAKULSRREGRETQAQIMVACTGESFQ	259		
QY	121	SHGGPEAPARLLVVVTTDSSHGGEELPAALKACEAGRVRYGIVLGHYLRQRDPSSFL	180		
DB	260	SHGGPEAPARLLVVVTTDSSHGGEELPAALKACEAGRVRYGIVLGHYLRQRDPSSFL	319		
QY	181	REIRTIASDPDERFFNV 198			
DB	320	REIRTIASDPDERFFNV 337			
RESULT 2					
ID	AAB64657	standard; Protein; 1152 AA.			
XX	AAB64657;				
XX	22-MAR-2001	(first entry)			
XX	Human secreted protein BLAST search protein SEQ ID NO: 167.				

XX	Cytostatic; immunosuppressive; noctropic; neuroprotective; antiviral;
KM	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
KM	vulnerable; anticonvulsant; antibacterial; antitumor; antiparasitic;
KM	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM	neurological disease; infection; human; secreted protein.
OS	Homo sapiens.
XX	WO20077197-A1.
PN	21-DEC-2000.
PD	01-JUN-2000; 2000MO-US14934.
XX	11-JUN-1999; 99US-0138599.
PR	(HUMA-) HUMAN GENOME SCI INC.
PA	(ROSE/) ROSEN C A.
XX	Rosen CA, Ruben SM, Komatsoulis GA;
PI	WPI: 2001-032312/04.
XX	Isolated nucleic acid molecule encoding a human secreted protein is
PT	used in preventing, treating or ameliorating a medical condition -
XX	Disclosure; Page 543-546; 558pp; English.
PS	The invention relates to the isolation of genes AA32757-f32803 encoding
CC	the human secreted proteins AB64549-B64594. The sequence is a search
CC	result from a BLAST homology search. The genes and proteins are useful
CC	for preventing, ameliorating or treating medical conditions, e.g. by
CC	protein or gene therapy. The genes are isolated from a range of human
CC	tissues disclosed in the specification. The nucleic acids, proteins,
CC	antibodies and (anti)agonists are useful in the diagnosis, treatment
CC	and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
CC	other cancers of the adrenal gland, bone, bone marrow, breast,
CC	gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC	disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC	anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC	multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC	cardiovascular disorders such as myocardial ischaemia; (d) wound
CC	healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC	and (f) infectious diseases such as viral, bacterial, fungal and
CC	parasitic infections.
CC	Sequence 1152 AA;
SQ	
Query Match	100.0%; Score 1025; DB 22; Length 1152;
Blast Local Similarity	100.0%; Pred. No. 7.6e-109; Mismatches 0; Indels 0; Gaps 0
Match 198; Conservative	
DB	140 GICARVDSFPOGSLAETAPORCPTPMVVIVLDGSNSIYPWSEVOTFLRLRVGLKFLDP 60
OY	1 GICARVDSFPOGSLAETAPORCPTPMVVIVLDGSNSIYPWSEVOTFLRLRVGLKFLDP 60
DB	61 EOIYGVGLVOYESPYHEWSLDFRKEEVVRAAKLSRRREGRETQAQIMACTBGEQ 120
OY	200 EOIYGVGLVOYESPYHEWSLDFRKEEVVRAAKLSRRREGRETQAQIMACTBGEQ 259
DB	121 SHGGPEPAAALLVVVTDDSHDGEELPALAKACEGRATRYIAVLGHILRROPDSSFL 180
OY	260 SHGGPEPAAALLVVVTDDSHDGEELPALAKACEGRATRYIAVLGHILRROPDSSFL 319
DB	181 REIRTIASDPDERFFNV 198
OY	320 REIRTIASDPDERFFNV 337
DB	
RESULT 3	
AB64658	
ID	AB64658 standard; Protein; 1152 AA.

XX AAB64658;
AC 22-MAR-2001 (first entry)
DT Human secreted protein BLAST search protein SEQ ID NO: 168.
XX
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anticancer;
XX vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein.
XX
XX Homo sapiens.
XX WO200071197-A1.
XX
XX 21-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US14934.
XX
XX 11-JUN-1999; 99US-0138599.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-032312/04.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition -
XX
XX
XX Disclosure; Page 547-551, 558pp; English.
XX
XX The invention relates to the isolation of genes AAF32757-F32803 encoding
XX the human secreted proteins AAB64549-864594. The sequence is used as a
XX query sequence for doing BLASTX searches to identify homologous
XX sequences. The genes and proteins are useful for preventing,
XX ameliorating or treating medical conditions, e.g. by protein or gene
XX therapy. The genes are isolated from a range of human tissues disclosed
XX in the specification. The nucleic acids, proteins, antibodies and
XX (ant)agonists are useful in the diagnosis, treatment and prevention of:
XX (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
XX adrenal gland, bone, bone marrow, breast, gastrointestinal
XX tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
XX disorders such as myocardial ischaemias; (d) wound healing; (e)
XX neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections.
XX
XX Sequence 1152 AA;
SQ

Query Match 100.0%; Score 1025; DB 22; Length 1152;
Best Local Similarity 100.0%; Pred. No. 7.6e-109;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDSFPOGSLAPTAQRCPITYMDVIVLDGNSNIYPMSEVOTFLRLVGLFIDP 60
DB 140 GICARVDSFPOGSLAPTAQRCPITYMDVIVLDGNSNIYPMSEVOTFLRLVGLFIDP 199
QY 61 EGIQVGLVOYGSPPHEWLSGPFRTKEEVRAAKNLRSREGSETTAQIMWACTBEGFSQ 120
DB 200 EGIQVGLVOYGSPPHEWLSGPFRTKEEVRAAKNLRSREGSETTAQIMWACTBEGFSQ 259
QY 121 SHGGRPEARLIVVTDSHGEELPALKACEAGRVRYGIAYVGHILRQRDPSSFL 180
DB 260 SHGGRPEARLIVVTDSHGEELPALKACEAGRVRYGIAYVGHILRQRDPSSFL 319
QY 181 REIRTIASDPDERFFENV 198

DB 320 REIRTIASDPDERFFENV 337

RESULT 4
ID AAY32242
ID AAY32242 standard; Protein; 1167 AA.
XX
XX AAY32242;
AC
XX
XX 15-FEB-2000 (first entry)
DT
XX
XX Human integrin subunit alpha-10.
DE
XX
XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
XX osteoarthritis; osteoarthritis; cancer; atherosclerosis;
XX inflammation; therapy; cartilage; chondrocyte; osteoblast;
XX fibroblast; vaccine; marker.
XX
XX Homo sapiens.
OS
XX
XX
XX Key Location/Qualifiers
XX
XX Peptide 1..22
XX /note= "signal peptide"
XX
XX Protein 23..1145
XX /note= "mature protein"
XX
XX Domain 23..1120
XX /note= "extracellular domain"
XX
XX Domain 1121..1145
XX /note= "transmembrane domain"
XX
XX Domain 1122..1167
XX /note= "cytoplasmic domain, specifically claimed
XX in Claim 21"
XX
XX Domain 162..359
XX /note= "I-domain"
XX
XX Binding-site 494..502
XX /note= "cation binding site motif"
XX
XX Binding-site 558..566
XX /note= "cation binding site motif"
XX
XX Binding-site 620..628
XX /note= "cation binding site motif"
XX
XX Modified-site 98
XX /note= "N-glycosylated"
XX
XX Modified-site 336
XX /note= "N-glycosylated"
XX
XX Modified-site 364
XX /note= "N-glycosylated"
XX
XX Modified-site 733
XX /note= "N-glycosylated"
XX
XX Modified-site 839
XX /note= "N-glycosylated"
XX
XX Modified-site 921
XX /note= "N-glycosylated"
XX
XX Modified-site 1018
XX /note= "N-glycosylated"
XX
XX Modified-site 1039
XX /note= "N-glycosylated"
XX
XX Modified-site /note= "N-glycosylated"
XX
XX W09951639-A1.
XX
XX 14-OCT-1999.
XX
XX 31-MAR-1999; 99WO-S000544.
XX
XX
XX 02-APR-1998; 98SE-0001164.
XX 28-JAN-1999; 99SE-0000319.
XX
XX (ACTI-) ACTIVE BIOTECH AB.
XX
XX Lundgren-Akerlund E;
XX
XX WPI; 2000-052639/04.
XX
XX N-PSDB; AA234719.

XX New isolated integrin subunit alpha-10, used as a marker or target
PT molecule for cells during development, regeneration and pathological
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT inflammation -

XX Claim 1; Fig 6; 90pp; English.

XX This sequence represents novel human chondrocyte integrin subunit
CC alpha-10 (Isa10). A splice variant is given in AAY32243. The
CC invention relates to a recombinant or isolated integrin heterodimer
CC comprising the alpha10 subunit in association with subunit beta
CC (especially beta-1). The heterodimer and the subunit alpha-10 can
CC be used as markers or targets of all types of cells, e.g. of
CC chondrocytes, osteoblasts and fibroblasts. They can also be used:
CC for treating pathological conditions involving Isa10, such as
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
CC for detecting the formation of cartilage during embryonal
CC development, physiological or therapeutic repair of cartilage,
CC or detecting regeneration of cartilage or chondrocytes during
CC transplantation of cartilage or chondrocytes; for selection and
CC analysis or for sorting, isolating or purification of chondrocytes
CC and for in vitro studies of differentiation of chondrocytes; and as
CC a target for anti-adhesive drugs or molecules in tendon, ligament,
CC skeletal muscle or other tissues where adhesion impairs the function
CC of the tissue (all claimed). Isa10 binding entities can be used to
CC determine the differentiation state of cells during embryonic
CC development, angiogenesis or development of cancer, in pathological
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
CC in tissue regeneration or in therapeutic and physiological repair
CC of cartilage (claimed). A vaccine comprising the integrin
CC heterodimer or subunit alpha-10 is also claimed. Isa10
CC polynucleotides, vectors, host cells and methods of producing
CC recombinant Isa10 are also claimed.

XX Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 21; Length 1167;
Best Local Similarity 100.0%; Pred. No. 7,7e-109;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVKGKLFIDP 60
DB 140 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVKGKLFIDP 199
QY 61 EOIQVGLVOYGESPVHEMSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWACTEGFSQ 120
DB 200 EOIQVGLVOYGESPVHEMSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWACTEGFSQ 259
QY 121 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTRYGIAVGLHRLRRORDPSSFL 180
DB 260 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTRYGIAVGLHRLRRORDPSSFL 319
QY 181 REIRTIASDPDERFFENV 198
DB -320 REIRTIASDPDERFFENV 337

RESULT 5

AAB64584 standard; Protein; 1167 AA.

AAB64584;

22-MAR-2001 (first entry)

Human secreted protein #37.

Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
neurological disease; infection; human; secreted protein.

XX Homo sapiens.
OS
PN WO20007197-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US14934.

XX 11-JUN-1999; 99US-0138599.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsu GJ GA;

XX WPI: 2001-032312/04.

XX N-PSDB; AAF32793.

PS Claim 11; Page 496-500; 558pp; English.

XX Sequences AAB64549-B64594 represent the amino acid sequences of 47
CC human secreted proteins encoded by the genes AAF32757-F32803. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic
CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

XX Sequence 1167 AA;

Query Match 100.0%; Score 1025; DB 22; Length 1167;
Best Local Similarity 100.0%; Pred. No. 7,7e-109;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVKGKLFIDP 60
DB 140 GICARVDASFPQGSIAPTAQRCPPTMDVIVLDGNSIYPMSEVQFLRLVKGKLFIDP 199
QY 61 EOIQVGLVOYGESPVHEMSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWACTEGFSQ 120
DB 200 EOIQVGLVOYGESPVHEMSLGDPRTEKEVVAARKNLSREGRRTKTAQAIWACTEGFSQ 259
QY 121 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTRYGIAVGLHRLRRORDPSSFL 180
DB 260 SHGGRPEARARLLVVTVDGSHDGEELPALAKACEAGRVTRYGIAVGLHRLRRORDPSSFL 319
QY 181 REIRTIASDPDERFFENV 198
DB 320 REIRTIASDPDERFFENV 337

RESULT 6

AAU76853 standard; Protein; 195 AA.

AAU76853;

21-MAY-2002 (first entry)

Human integrin alpha subunit Alpha 10 A domain.

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KW ischaemia-reperfusion injury; immune complex; parasitic disease;
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200209737-A1.
 XX PD 07-FEB-2002.
 XX PF 31-JUL-2001; 2001WO-US23957.
 XX PR 31-JUL-2000; 2000US-221950P.
 XX PR 11-JAN-2001; 2001US-0758493.
 XX PR 13-MAR-2001; 2001US-0805354.
 XX PA (GHEO) GEN HOSPITAL CORP.
 XX PI Arnaout AM, Li R, Xiong J;
 XX DR WPI; 2002-188687/24.
 XX PT Novel high affinity integrin polypeptide useful for treating restenosis
 PT and parasitic diseases, comprises all or part of variant integrin alpha
 PT subunit A domain or variant integrin beta subunit A-like domain -
 XX
 XX PS Example 2; Fig 5; 55pp; English.
 XX CC The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD1b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC the human integrin alpha subunit Alpha 10 A domain.
 XX
 XX SQ Sequence 195 AA;
 XX
 XX Query Match 89.0%; Score 912; DB 23; Length 195;
 XX Best Local Similarity 100.0%; Pred. No. 6.5e-97;
 XX Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQGESPVHMSLCD 82
 DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQGESPVHMSLCD 60
 QY 83 FRTKEVVAANKLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGESH 142
 DB 61 FRTKEVVAANKLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGESH 120
 QY 143 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 198
 DB 121 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 176

DE Human integrin alpha subunit Alpha 10 variant A domain.
 XX
 XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
 KW mutein.
 XX OS Homo sapiens.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Misc-difference 193
 FT /note= "wild-type Ile substituted by any other amino
 FT acid"
 XX
 XX PN WO200209737-A1.
 XX PD 07-FEB-2002.
 XX PF 31-JUL-2001; 2001WO-US23957.
 XX PR 31-JUL-2000; 2000US-221950P.
 XX PR 11-JAN-2001; 2001US-0758493.
 XX PR 13-MAR-2001; 2001US-0805354.
 XX PA (GHEO) GEN HOSPITAL CORP.
 XX PI Arnaout AM, Li R, Xiong J;
 XX DR WPI; 2002-188687/24.
 XX PT Novel high affinity integrin polypeptide useful for treating restenosis
 PT and parasitic diseases, comprises all or part of variant integrin alpha
 PT subunit A domain or variant integrin beta subunit A-like domain -
 XX
 XX PS Claim 53; Page -; 55pp; English.
 XX CC The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD1b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC a human integrin alpha subunit Alpha 10 variant A domain.
 CC Note: This variant sequence is not featured in the specification but has
 CC been derived from the wild-type protein shown in AAU76853.
 XX
 XX SQ Sequence 195 AA;
 XX
 XX Query Match 89.0%; Score 912; DB 23; Length 195;
 XX Best Local Similarity 100.0%; Pred. No. 6.5e-97;
 XX Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQGESPVHMSLCD 82
 DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIQVGLVQGESPVHMSLCD 60
 QY 83 FRTKEVVAANKLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGESH 142
 DB 61 FRTKEVVAANKLSRRREGRETKTAQAIMVACTEGFSQSHGGRPEAARLLVVTGESH 120
 QY 143 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 198
 DB 121 GELPAAKACAGAVTRYGIAGVILGHYLRORDPSSFLREIRTIASDDPERFFNV 176

RESULT 8
 ID AAY32282
 XX AAY32282 standard; Protein; 303 AA.
 AC AAY32282;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Mouse integrin subunit alpha-10 partial sequence.
 XX
 KW Integrin alpha-10; Itsa10; mouse; trauma; rheumatoid arthritis;
 XX osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 XX fibroblast; vaccine; marker.
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 15
 FT /note="encoded by TWG"
 FT
 XX
 PN WO951639-A1.
 XX
 PD 14-OCT-1999.
 XX
 PF 31-MAR-1999; 99WO-SE00544.
 XX
 PR 02-APR-1998; 98SE-0001164.
 PR 28-JAN-1999; 99SE-0000319.
 XX
 XX
 PA (ACT1-) ACTIVE BIOTECH AB.
 XX
 PI Lundgren-Akerlund E;
 XX
 DR WPI; 2000-052639/04.
 DR N-PSDB; AA34835.
 XX
 XX
 PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation -
 PT
 XX
 PS Example 15; Fig 15; 90pp; English.
 XX
 CC This sequence represents a polypeptide encoded by an isolated
 CC partial genomic DNA (see AA34835) of murine integrin subunit
 CC alpha-10 (Itsa10). The mouse genomic alpha-10 sequence was used
 CC to generate a targeting vector for knockout experiments. The
 CC invention relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can
 CC be used as markers or targets of all types of cells, e.g. of
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used
 CC for treating pathological conditions involving Itsa10, such as:
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage,
 CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). Binding entities, polynucleotides
 CC (PNS) or oligonucleotides (ONS) can be used for e.g. determining the
 CC differentiation-state of cells during embryonic development,
 CC angiogenesis or development of cancer, in pathological conditions
 CC such as rheumatoid arthritis, osteoarthritis or cancer, in tissue
 CC regeneration or in therapeutic and physiological repair of
 CC cartilage (claimed). The PNS and ONS can also be used for detecting
 CC pathological conditions such as atherosclerosis or inflammation

CC (claimed). A vaccine comprising the integrin heterodimer or
 CC subunit alpha-10 is also claimed. Vectors, host cells and methods
 CC of producing recombinant Itsa10 are provided.
 XX
 SQ Sequence 303 AA;
 XX
 Query Match 77.8%; Score 797.5; DB 21; Length 303;
 Best Local Similarity 95.1%; Pred. No. 2.2e-83;
 Matches 156; Conservative 5; Mismatches 2; Indels 1; Gaps 1;
 XX
 QY 1 GTCARVDASFPQGSGLAFTAQRCPTVMVIVLDGSSNITPMSEVQTLRLVGLFTDP 60
 DB 141 GTCARVDASFPQGSGLAFTAQRCPTVMVIVLDGSSNITPMSEVQTLRLVGLFTDP 200
 QY 61 EGIQVGLVQYGEPPVHEWSLGDPTFKEEVRAKXLSREGRETKTAQAIWACTEGFSQ 120
 DB 201 EGIQVGLVQYGEPPVHEWSLGDPTFKEEVRAKXLSREGRETKTAQAIWACTEGFSQ 259
 QY 121 SHGGRPEARLLVVTDESHDGEELPALAKACEAGRVTRYGIA 164
 DB 260 SHGGRPEARLLVVTDESHDGEELPALAKACEAGRVTRYGIA 303
 XX
 RESULT 9
 ID AAB50087
 XX AAB50087 standard; Protein; 1188 AA.
 XX
 AC AAB50087;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Murine A259.
 XX
 KW Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
 KW kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
 KW rheumatoid arthritis.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= Signal_peptide
 FT 23..1188
 FT /label= Mature_protein
 FT 1..1141
 FT /label= Extracellular_domain
 FT 39..74
 FT /label= Integrin_alphasubunit_repeat_domain_#1
 FT 115..157
 FT /label= Integrin_alphasubunit_repeat_domain_#2
 FT 164..345
 FT /label= I_domain
 FT 367..392
 FT /label= Integrin_alphasubunit_repeat_domain_#3
 FT 421..455
 FT /label= Integrin_alphasubunit_repeat_domain_#4
 FT 478..516
 FT /label= Integrin_alphasubunit_repeat_domain_#5
 FT 540..575
 FT /label= Integrin_alphasubunit_repeat_domain_#6
 FT 602..640
 FT /label= Integrin_alphasubunit_repeat_domain_#7
 FT 1142..1164
 FT /label= Transmembrane_domain
 FT 1165..1188
 FT /label= Cytoplasmic_domain
 XX
 WO200073339-A1.
 XX
 PD 07-DEC-2000.
 XX
 PF 15-MAY-2000; 2000WO-US13262.
 XX


```

PR 28-MAY-1999; 99US-0322790.
PR 27-APR-2000; 2000US-0561263.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lora JM;
XX
XX WPI; 2001-041142/05.
XX N-PSDB; AAC91904, AAC91905.
XX
XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
XX diagnosis of fibrosis, e.g. of the liver
XX
XX Claim 8; Fig 5; 164pp; English.
XX
XX The present sequence is murine integrin alpha subunit, A259. A259 is
XX homologous with the alpha and alpha2 integrin subunits and is
XX overexpressed in fibrosis. A259 is implicated in regulation of
XX proliferation, differentiation and/or function of many different cell
XX types. Inhibitors of A259 activity are useful for the treatment of liver
XX disease, particularly fibrosis, and also fibrosis in other organs
XX (specifically lung and kidney). In addition, A259 can be used for
XX treatment and prevention of cancer, osteoporosis, acute myeloid
XX leukaemia, HIV infection, and rheumatoid arthritis.
XX
XX Sequence 1188 AA;
SQ
Query Match 53.4%; Score 547; DB 23; Length 1188;
Best Local Similarity 53.0%; Pred. No. 1.5e-53;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;
QY 1 GICARVDASFOPGSLAPTAQRCPTMDVIVLDGNSIYPWSEVQFLRLVGLKFLDP 60
DB 127 GMCGRVNSNFRSKTVAPALQRCQTYMDIVLDGNSIYPWVEVQHFLINLKKFYIGP 196
QY 61 EGIQVGLVOYGSPVHEWLSGPFRTKEEVNRAKLSRREGSEPTKTAQIMVACTEGESQ 120
DB 197 GGIQVGIVOYGEDAVHEFLINDYRSKVDVEAASHIEQRCGETRTAFGEIPARESEAFQK 256
QY 121 SHGSGPEARLILWVTDDGSHDGEELPALKACAGRVTRGIVLGHYLRQRQPSSTL 180
DB 257 --GGRKGAKKWIVITDGHSDSPDEKVIROSEKDNVTRVAVVAGYNNRGINPEFTL 314
QY 181 REIRTIASDPDERFFPNV 198
DB 315 NEIKYIASDPDDKHFFNV 332
QY
DB
RESULT 10
AAU10552
ID AAU10552 standard; Protein; 1188 AA.
XX
XX AAU10552;
XX
XX 14-FEB-2002 (first entry)
XX
XX Murine A259 polypeptide.
XX
XX Mouse; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
XX liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
XX cartilage associated disorder; haematopoietic disorder; bone marrow;
XX immune related disease; apoptotic disorder; neuronal tissue disease;
XX neurodegenerative disease; gene therapy; cancer; cytostatic; osteoparh;
XX nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
XX antidiabetic; antianaemic; antiallergic; antiaesthetic; dermatological;
XX antidiabetic; anticonvulsant; antiparkinsonian.
XX
XX Mus musculus.
XX
XX Key Location/Qualifiers
XX Peptide 1..22
XX FT /note= "Signal peptide"
XX Domain 1..1141

```

```

FT Protein /note= "Extracellular domain"
FT 23..1188 /note= "Murine murine A259"
FT 39..74 /note= "Integrin alpha repeat domain"
FT 115..157 /note= "Integrin alpha repeat domain"
FT 164..345 /note= "Integrin alpha repeat domain"
FT 367..392 /note= "I domain or Von Willebrand Factor type A domain"
FT 421..455 /note= "Integrin alpha repeat domain"
FT 478..516 /note= "Integrin alpha repeat domain"
FT 540..575 /note= "Integrin alpha repeat domain"
FT 602..640 /note= "Integrin alpha repeat domain"
FT 1142..1164 /note= "Integrin alpha repeat domain"
FT 1165..1188 /note= "Transmembrane domain"
FT 1165..1188 /note= "Cytoplasmic domain"
XX
XX WO200181414-A2.
XX
XX 01-NOV-2001.
XX
XX 27-APR-2001; 2001WO-US13516.
XX
XX 27-APR-2000; 2000US-0561263.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lora J;
XX
XX WPI; 2002-041397/05.
XX N-PSDB; AAS16874.
XX
XX New A259 nucleic acids and polypeptides, which comprise integrin alpha
XX subunit, useful for diagnosing, preventing or treating e.g. liver
XX disease, kidney or lung fibrosis, cancers, blood disorders or immune
XX related diseases
XX
XX Claim 9; Fig 5; 168pp; English.
XX
XX The invention relates to human and murine A259 nucleic acid molecules
XX which encode secreted proteins with homology to integrin alpha subunits,
XX specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
XX are useful for treating liver disease or fibrosis, particularly kidney
XX fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
XX useful for diagnosing, preventing or treating cartilage and bone
XX associated disorders (such as bone cancer, achondroplasia, myeloma,
XX fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
XX osteoporosis), bone marrow, blood and haematopoietic disorders (such as
XX acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune
XX related diseases (such as HIV, viral infections, cancers, T cell
XX autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
XX asthma and psoriasis), apoptotic disorders (such as systemic lupus
XX erythematosus and insulin-dependent diabetes mellitus), diseases of the
XX neuronal tissues (such as epilepsy and muscular dystrophy) and
XX neurodegenerative diseases (such as Parkinson's disease and Huntington's
XX disease). This sequence represents the murine A259 polypeptide.
XX
XX Sequence 1188 AA;
SQ
Query Match 53.4%; Score 547; DB 23; Length 1188;
Best Local Similarity 53.0%; Pred. No. 1.5e-53;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;
QY 1 GICARVDASFOPGSLAPTAQRCPTMDVIVLDGNSIYPWSEVQFLRLVGLKFLDP 60
DB 137 GMCGRVNSNFRSKTVAPALQRCQTYMDIVLDGNSIYPWVEVQHFLINLKKFYIGP 196

```



```

PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236386P.
PR 29-SEP-2000; 2000US-236399P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX (ROSE/) ROSEN C A.
XX (RUBE/) RUBEN S M.
XX (BARA/) BARASH S C.
PI Rosen CA, Ruben SM, Barash SC;
XX WPI; 2002-470713/50.
DR N-PSDB; ABO66558.
XX
PT New nucleic acid encoding human proteins, useful for diagnosis,
PT treatment and prevention of e.g. osteoporosis, also related
PT polypeptides and antibodies
XX
PS Claim 11; SEQ ID NO 313; 235bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABO66521-ABO66785) and proteins
XX (ABP47846-ABP48110) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX disease, hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from USPTO at segdata.uspto.gov/sequence.html?DocID=999909764870.
XX
XX Sequence 707 AA:
XX
Query Match 53.1%; Score 544; DB 23; Length 707;
Best Local Similarity 53.0%; Pred. No. 1,5e-53;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;
XX
1 GICAVDASFOPOGSLAPTRACPTMYMDVIVLDGNSNTYPMSEVOTFLRLVGLFIDP 60
153 GMSCRNSNFRSKTVAPALQRCQYMDIVILVDSGNSNTYPMVEVQHFILNLLKFFYIGP 212
61 EDIOGLVGYGSPYHENGIDGPFRTKEEVRAAKNLSRREGRETKQAQIMVACTEGFSQ 120
213 GGIQGVGVGYGSDVHNEHILNDRSKVDVEASHIEQGGCTETRTAFIEFAABEAFQK 272
121 SHGAPPEARLLIVVTGDSHGEELPALKACENAGRTVRYGIAVLGHYLRQRDPSSFTL 180

```

```

DB 273 --GGRGAKVMIVITDGDHSDPDLKVIQOSERDNTRYAVAVLGYNRRGINPETL 330
OY 181 REIRTIASDPDERFFNV 198
DB 331 NEIKYIASDPDKPFNV 348
RESULT 13
AAB25590
ID AAB25590 standard; Protein; 1034 AA.
AC AAB25590;
DT 21-NOV-2000 (first entry)
XX
XX Protein encoded by human secreted protein gene #7 clone HOHBV69.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
XX anti-rheumatic; dermatological; antiproliferative; antiarteriosclerotic;
XX anticancer; vulnary; antiviral; antibacterial; antifungal;
XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
XX Crohn's disease; nephritis; hyperproliferative disorder;
XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
XX melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
XX Homo sapiens.
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US25031.
XX
XX 28-OCT-1998; 98US-0105971.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny J, Moore PA, Wei Y;
XX Greene JM;
XX
XX WPI; 2000-387742/33.
XX
XX Isolated nucleic acid molecules encoding human secreted proteins are
XX used for the prevention, amelioration and treatment of autoimmune,
XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases -
XX
XX Claim 1; Page 678-682; 803bp; English.
XX
XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given
XX in AAB80606-A80623 encode the 12 secreted protein sequences given in
XX AAB25576-825593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant;
XX anti-inflammatory; antiarthritic; antirheumatic; dermatological;
XX antiproliferative; antiarteriosclerotic; anticancer; vulnary;
XX antiviral; antibacterial; and antifungal activity. The proteins,
XX polypeptides, agonists and antagonists may be used to treat prevent
XX and/or diagnose various disease, disorders and conditions examples of
XX which include: immune disorders e.g. Addison's disease, rheumatoid
XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
XX hyperproliferative disorders such as paraproteinemia and purpura;
XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
XX sequences may also be used in wound healing and the treatment of
XX infectious diseases. The human secreted protein gene #7 and protein
XX sequences are represented in sequences AAB80612 and AAB25582. Secreted
XX protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
XX AAB80652-A80661 represent genes which are related to the secreted protein
XX gene#7.

```


DR N-PSDB; AAS22772.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX

PS Example 4; Page 828-831; 894pp; English.
 XX

CC The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/ elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in treating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.
 XX

SQ Sequence 1188 AA:

Query Match 53.1%; Score 544; DB 22; Length 1188;
 Best Local Similarity 53.0%; Pred. No. 3.3e-53;
 Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

QY 1 GTCARVDASFPQSGSLAPTAORCPTMDVYIVDGSNSIYPWSEVOTFLRLVGLKFLIDP 60
 Db 137 GMSRVNSNFRPSKTVAPALQRCQITMDVIVDGSNSIYPWSEVOTFLRLVGLKFLIDP 196
 QY 61 EQIOVGLVOYGESPVHEWSLGDFTKEEVRAKNLSRREGRETAKAIMVACTEGFSQ 120
 Db 197 GQIOVGLVOYGEVVAHEPHLNDYRSVKDYEAASHIEQGGTETRTAFGIBFARSEAFQK 256
 QY 121 SHQGRPEARLVLVVTDDGSHDGEELPAALKACEAGRVTRYGIATVGHYLRQRDPSSFL 180
 Db 257 --GGRGAKKVMIVIDGSHDSDPLEKVIQOGERDNTVRYAVAVLGYNRRGINDETFL 314
 QY 181 REIRTIASDPDERFFPNV 198
 Db 315 NEIKYIASDPDDKHFPNV 332

Search completed: July 16, 2003, 07:52:05
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 07:55:38 ; Search time 14.318 Seconds
(without alignments)
1639.577 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337
Perfect score: 1025
Sequence: 1 GICARVDSFQPGSLAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	89.0	195	US-09-805-354-7	Sequence 7, Appl1
2	912	89.0	195	US-10-144-259-7	Sequence 7, Appl1
3	544	53.1	707	US-10-125-540-313	Sequence 313, App
4	544	53.1	707	US-09-764-870-313	Sequence 313, App
5	544	53.1	1034	US-09-984-130-43	Sequence 43, Appl
6	544	53.1	1034	US-09-836-353A-43	Sequence 43, Appl
7	544	53.1	1189	US-09-984-130-35	Sequence 35, Appl
8	544	53.1	1189	US-09-836-353A-35	Sequence 35, Appl
9	492	48.0	193	US-09-805-354-8	Sequence 8, Appl1
10	492	48.0	193	US-10-144-259-8	Sequence 8, Appl1
11	467.5	45.6	1151	US-09-984-130-103	Sequence 103, App
12	467.5	45.6	1151	US-09-836-353A-103	Sequence 103, App
13	452.5	44.1	214	US-10-061-658-5	Sequence 5, Appl1
14	449.5	43.9	214	US-10-061-658-6	Sequence 6, Appl1
15	449.5	43.9	214	US-10-061-658-9	Sequence 9, Appl1
16	441	43.0	1181	US-10-160-354-2	Sequence 2, Appl1
17	433.5	42.3	212	US-09-996-738-5	Sequence 5, Appl1
18	430.5	42.0	212	US-09-996-738-6	Sequence 6, Appl1
19	427	41.7	195	US-09-805-354-5	Sequence 5, Appl1

20	427	41.7	195	US-10-144-259-5	Sequence 5, Appl1
21	399	38.9	195	US-09-805-354-6	Sequence 6, Appl1
22	399	38.9	195	US-10-144-259-6	Sequence 6, Appl1
23	285.5	27.9	148	US-10-125-540-284	Sequence 284, App
24	285.5	27.9	148	US-10-125-540-444	Sequence 444, App
25	285.5	27.9	148	US-09-764-870-284	Sequence 284, App
26	285.5	27.9	148	US-09-764-870-444	Sequence 444, App
27	282	27.5	103	US-10-177-293-472	Sequence 472, App
28	282	27.5	103	US-09-764-870-472	Sequence 472, App
29	210.5	20.5	3063	US-10-177-293-63	Sequence 63, Appl
30	209.5	20.4	3063	US-10-177-293-61	Sequence 61, Appl
31	198	19.3	413	US-09-891-943-101	Sequence 101, App
32	198	19.3	413	US-09-350-259-101	Sequence 101, App
33	191	18.6	776	US-10-000-512-8	Sequence 8, Appl1
34	191	18.6	915	US-09-905-291A-34	Sequence 34, Appl1
35	191	18.6	915	US-09-902-853-34	Sequence 34, Appl1
36	191	18.6	915	US-09-907-824-34	Sequence 34, Appl1
37	191	18.6	915	US-09-907-841-34	Sequence 34, Appl1
38	191	18.6	915	US-09-904-011-34	Sequence 34, Appl1
39	191	18.6	915	US-10-028-072-294	Sequence 294, App
40	191	18.6	915	US-09-906-742-34	Sequence 34, Appl1
41	191	18.6	915	US-10-121-049-294	Sequence 294, App
42	191	18.6	915	US-10-123-904-294	Sequence 294, App
43	191	18.6	915	US-10-140-470-294	Sequence 294, App
44	191	18.6	915	US-09-906-838-34	Sequence 34, Appl1
45	191	18.6	915	US-09-907-613-34	Sequence 34, Appl1

ALIGNMENTS

```
RESULT 1
US-09-805-354-7
; Sequence 7, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US/09/805,354
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-7

Query Match      89.0%; Score 912; DB 9; Length 195;
Best Local Similarity 100.0%; Pred. No. 1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      23 CPTVMDVIVDGSNSIYPMSEVOTFLRLVYKGLFIDEQIOVGLVOYGESPVHWSIGCD 82
DB      1 CPTVMDVIVDGSNSIYPMSEVOTFLRLVYKGLFIDEQIOVGLVOYGESPVHWSIGCD 60
QY      83 FRTREYVRAKNLSRRREGRETAKQAIVMVCCTGFSQSGRPPAAALLVVVTGDSHD 142
DB      61 FRTREYVRAKNLSRRREGRETAKQAIVMVCCTGFSQSGRPPAAALLVVVTGDSHD 120
QY      143 GEEUPAALKACAGVTRVYGLAVLGHYLRORDSSFLREIRTIASDPDERFFENV 198
DB      121 GEEUPAALKACAGVTRVYGLAVLGHYLRORDSSFLREIRTIASDPDERFFENV 176

RESULT 2
```


US-09-984-130-43

Query Match	53.1%;	Score 544;	DB 9;	Length 1034;
Best Local Similarity	53.0%;	Pred. NO. 2.7e-49;		
Matches 105;	Conservative 39;	Mismatches 52;	Indels 2;	Gaps 1

[illegible]

RESULT 6
US-09-836-353A-43

Query Match	53.1%	Score 544	DB 9	Length 1034
Best Local Similarity	53.0%	Pred. NO. 2.7e-49		
Matches 105, Conservative	39	Mismatches 52	Indels 2	Gaps 1

[illegible]

```

RESULT 7
US-09-984-130-35
; Sequence 35, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-984-130-35

Query Match          53.1%; Score 544; DB 9; Length 1189;
Best Local Similarity 53.0%; Pred. No. 3,36-49;
Matches 105; Conservative 39; Mismatches 52; Indels 2; Gaps 1;

```

Oy		G I A R D A S F O G S L A P T A C R P Y M V N V L D G S N S I Y W S V O F E L R A I N K G J I D P	60
Dd	137	G M O S R A N F R S K S T Y A P A L O C Q T I M V I A L D G S N S I Y W E V O F L I N I L K K Y I G P	196
Oy	61	E Q I O V G L V O Y G E S P V H E M S L I D F R T K E E V N A A N K L S R R E G E T K T A Q A I M V A C T E G S Q	120
Dd	197	G Q I Q V G V O Y G E D V V H E F L N D Y R S V K U V E A S H I B E Q R G T E R T A F G I E P A S E A F O X	256
Oy	121	S H O G R E P A A R L V V T D G S H D G E E L P A L K A C E A G R Y T R G I A V L G H Y I R R O R D P S S F L	180
Dd	257	--GGRGAACKWMI VTLDGESHDSPLEKVIQOESERDVNRYAVALGYNNRRGINPETFL	314
Oy	181	R E I R T I A S D D E R R F F N V	198
Dd	315	N E I K Y I A S D P D K G H F F N V	332

```

US-09-836-353A-35
RESULT 8
US-09-836-353A-35
; Sequence 35, Application US/0986353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PFA69P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-35

```

Query Match	53.1%	Score 544;	DB 9;	Length 1189;
Best Local Similarity	53.0%;	Pred. No. 3,36-49;		
Matches 105;	Conservative 39;	Mismatches 52;	Indels 2;	Gaps 1;
QY	1	GTCARVVASFPQSSSLPTAQRCTPYNDVIVLVDGNSISYPMSEVQTFRLRLVQKLFDP	60	

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8
Db 137 GMSRVNSNFRFSKTVAPALQRCQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYIGP 196
QY 61 EQIQVGLVQGESPVHEWSLGDFTKEEVRAAKNLSRREGRETKTQAQIIVACTGFSQ 120
Db 197 GQIQVGVVQGEDVHEFHLNDYRSYDVVEASHLEQRCGTETRTAFGIEFARSEAFQK 256
QY 121 SHGGRPEARLLVVTVDGESHDEELPAALACAGAVTRFYGIAYLGHYLRORDPSSFL 180
Db 257 --GGRGAKKVMIVITDGSHEHSDPLEKVIQOSEBDNVTYVAVALGYNNRGINPFTL 314
QY 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 9
US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaud, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; PRIOR FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-8

Query Match 48.0%; Score 492; DB 9; Length 193;
Best Local Similarity 54.5%; Pred. No. 9.8e-45;
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;

QY 23 CPTVMDVIVLDGNSIYPMWEVQTFRLRVGLFLDPQIQVGLVQGESPVHEWSLGD 82
Db 1 CQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYIGQIQVGVQGEDVHEFHLND 60
QY 83 FRTKEEVRAAKNLSRREGRETKTQAQIIVACTGFSQSHGGRPEARLLVVTVDGESH 142
Db 61 YRSVKDVEASHLEQRCGTETRTAFGIEFARSEAFQK--GGRGAKKVMIVITDGS 118
QY 143 GELPAAKACAGAVTRFYGIAYLGHYLRORDPSSFLREIRTIASDPDERFFNV 198
Db 119 SPLEKVIQOSEBDNVTYVAVALGYNNRGINPFTLEIKYIASDPDDKHFFNV 174

RESULT 10
US-10-144-259-8
; Sequence 8, Application US/10144259
; Publication No. US20030109691A1
; GENERAL INFORMATION:
; APPLICANT: Arnaud, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-548001
; CURRENT APPLICATION NUMBER: US/10/144,259
; PRIOR FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
```

```

; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-259-8
Db 137 GMSRVNSNFRFSKTVAPALQRCQTYMDIYIVLDGNSIYPMWEVQHFLINILKKFYIGP 196
QY 61 EQIQVGLVQGESPVHEWSLGDFTKEEVRAAKNLSRREGRETKTQAQIIVACTGFSQ 120
Db 197 GQIQVGVVQGEDVHEFHLNDYRSYDVVEASHLEQRCGTETRTAFGIEFARSEAFQK 256
QY 121 SHGGRPEARLLVVTVDGESHDEELPAALACAGAVTRFYGIAYLGHYLRORDPSSFL 180
Db 257 --GGRGAKKVMIVITDGSHEHSDPLEKVIQOSEBDNVTYVAVALGYNNRGINPFTL 314
QY 181 REIRTIASDPDERFFNV 198
Db 315 NEIKYIASDPDDKHFFNV 332

RESULT 11
US-09-984-130-103
; Sequence 103, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P2
; CURRENT APPLICATION NUMBER: US/09/984,130
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-103

Query Match 45.6%; Score 467.5; DB 9; Length 1151;
Best Local Similarity 46.5%; Pred. No. 4.9e-41;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

QY 1 GICARVDSAPQGSIAPTAQCPTVMDVIVLDGNSIYPMWEVQTFRLRVGLFLDP 60
Db 118 GICSDVSPFPVYVNSIAP-VEGCTQDVIYIVLDGNSIYPMWDSVTALNDLKRMDIGP 176
QY 61 EQIQVGLVQGESPVHEWSLGDFTKEEVRAAKNLSRREGRETKTQAQIIVACTGFSQ 120
Db 177 KQIQVGIQYGENVTHEENLNKYSTEVVLAKKIYQRGGRQMTALGTATKAEATFE 236
QY 121 SHGGRPEARLLVVTVDGESHDEELPAALACAGAVTRFYGIAYLGHYLRORDPSSFL 180
Db 237 ARGARRGVKVMIVITDGSHEHSDHNLKVKVIQDCEDENIQRFSIALGSGYNNRGINLSTEFV 296
QY 181 REIRTIASDPDERFFNV 198
Db 297 EIKSIASEPTEKHFNV 314
```

RESULT 12
US-09-836-353A-103
; Sequence 103, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PF489P1
; CURRENT APPLICATION NUMBER: US/09/836,353A
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 103
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-353A-103

Query Match 45.6%; Score 467.5; DB 9; Length 1151;
Best Local Similarity 46.5%; Pred. No. 4.9e-41;
Matches 92; Conservative 36; Mismatches 69; Indels 1; Gaps 1;

QY 1 GICAVDASFPQGSIAPTAORCPITYMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOV 60
DB 118 GICSVSPFPQVNSIAP-VQECSTQLDIVYLDGNSIYPMDSVTAFINDLKRMDIGP 176
QY 61 EGIQVGLVQYGSPPHEMSIGDPRTKEEVVRAXKLSRREGRETKQAQIMVACTEGFSQ 120
DB 177 KOTQVGIQYQENVTHERFNKYSSTEEVLVAKKIVQRGSGQMTALCTDTRKAEAFTE 236
QY 121 SHGGEPEARLLVVTDSGSHDGEELPAALKACEAGRVRYGIAVGHYLRQRDPSSFL 180
DB 237 ARGARGVYKVVIVTDSGSHDNHRLKKVYIQDCEDENIQRFSAIIGSYNRGNLSTKEKV 296
QY 181 REIRTIASDPDERFFPNV 198
DB 297 EIKSIASEPTEKHFFNV 314

RESULT 13
US-10-061-658-5
; Sequence 5, Application US/10061658
; Publication No. US20020182213A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Rat
US-10-061-658-5

Query Match 44.1%; Score 457.5; DB 9; Length 214;
Best Local Similarity 46.6%; Pred. No. 1.9e-40;
Matches 90; Conservative 34; Mismatches 68; Indels 1; Gaps 1;

QY 6 VDASFQPGSLAPTAORCPITYMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOV 65
DB 1 VSPTFQVNSIAP-VQECSTQLDIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOV 59
QY 66 GLVQYGSPPHEMSIGDPRTKEEVVRAXKLSRREGRETKQAQIMVACTEGFSQSHGR 125
DB 60 GIVQYGENVTHERFNKYSSTEEVLVAKKIVQRGSGQMTALCTDTRKAEAFTEARGAR 119
QY 126 PEARLLVVTDSGSHDGEELPAALKACEAGRVRYGIAVGHYLRQRDPSSFLREIRT 185
DB 120 RGVKKVVIVTDSGSHDNHRLKKVYIQDCEDENIQRFSAIIGSYNRGNLSTKEKVEEIKS 179
QY 186 IASDPDERFFPNV 198
DB 180 IASEPTEKHFFNV 192

RESULT 14
US-10-061-658-6
; Sequence 6, Application US/10061658
; Publication No. US20020182213A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US
; CURRENT APPLICATION NUMBER: US/10/061,658
; CURRENT FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/137,214
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: 60/130,847
; PRIOR FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-061-658-6

Query Match 43.9%; Score 449.5; DB 9; Length 214;
Best Local Similarity 46.1%; Pred. No. 4e-40;
Matches 89; Conservative 35; Mismatches 68; Indels 1; Gaps 1;

QY 6 VDASFQPGSLAPTAORCPITYMDVIVYLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOV 65
DB 1 VSPTFQVNSIAP-VQECSTQLDIVYLDGNSIYPMDSVTAFINDLKRMDIGPKQIOV 59
QY 66 GLVQYGSPPHEMSIGDPRTKEEVVRAXKLSRREGRETKQAQIMVACTEGFSQSHGR 125
DB 60 GIVQYGENVTHERFNKYSSTEEVLVAKKIVQRGSGQMTALCTDTRKAEAFTEARGAR 119
QY 126 PEARLLVVTDSGSHDGEELPAALKACEAGRVRYGIAVGHYLRQRDPSSFLREIRT 185
DB 120 RGVKKVVIVTDSGSHDNHRLKKVYIQDCEDENIQRFSAIIGSYNRGNLSTKEKVEEIKS 179
QY 186 IASDPDERFFPNV 198
DB 180 IASEPTEKHFFNV 192

RESULT 15
US-10-061-658-9
; Sequence 9, Application US/10061658
; Publication No. US20020182213A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Gotwals, Philip
; APPLICANT: Koteliansky, Victor
; TITLE OF INVENTION: Method for the Treatment of Fibrosis
; FILE REFERENCE: A073US

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 7.79747 Seconds
(without alignments)
747.132 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025
Sequence: 1 GICARVDSFQPGSLAPTA.....FLREIRTIASDPDERFFPNV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	205	20.0	496	1	US-08-462-128-37 Sequence 37, Appl
2	205	20.0	496	1	US-08-463-180-37 Sequence 37, Appl
3	205	20.0	496	2	US-08-001-078A-1 Sequence 1, Appl
4	205	20.0	496	2	US-08-897-443-4 Sequence 4, Appl
5	205	20.0	496	2	US-08-463-218-1 Sequence 1, Appl
6	205	20.0	496	5	PCT-US94-00253-1 Sequence 1, Appl
7	198	19.3	413	1	US-08-485-618-101 Sequence 101, App
8	198	19.3	413	2	US-08-605-672-101 Sequence 101, App
9	198	19.3	413	2	US-08-482-293A-101 Sequence 101, App
10	198	19.3	413	2	US-08-943-363-101 Sequence 101, App
11	198	19.3	413	4	US-09-193-043-101 Sequence 101, App
12	198	19.3	413	4	US-09-688-307A-101 Sequence 101, App
13	196	19.1	956	2	US-08-897-443-3 Sequence 3, Appl
14	191	18.6	638	2	US-08-897-443-1 Sequence 2, Appl
15	177.5	17.3	1178	3	US-08-199-776-2 Sequence 2, Appl
16	177.5	17.3	1178	3	US-08-663-731-2 Sequence 2, Appl
17	177.5	17.3	1178	3	US-08-879-338-2 Sequence 2, Appl
18	177.5	17.3	1178	5	PCT-US95-02044-2 Sequence 2, Appl
19	177.5	17.3	1179	4	US-09-293-238B-2 Sequence 2, Appl
20	169	16.5	1155	1	US-08-286-889-46 Sequence 46, Appl
21	169	16.5	1155	1	US-08-485-618-46 Sequence 46, Appl
22	169	16.5	1155	1	US-08-362-652-46 Sequence 46, Appl
23	169	16.5	1155	2	US-08-605-672-46 Sequence 46, Appl
24	169	16.5	1155	2	US-08-482-293A-46 Sequence 46, Appl
25	169	16.5	1155	2	US-08-943-363-46 Sequence 46, Appl
26	169	16.5	1155	4	US-09-193-043-46 Sequence 46, Appl
27	169	16.5	1155	4	US-09-688-307A-46 Sequence 46, Appl

28	169	16.5	1161	1	US-08-485-618-53 Sequence 53, Appl
29	169	16.5	1161	2	US-08-362-652-53 Sequence 53, Appl
30	169	16.5	1161	2	US-08-605-672-53 Sequence 53, Appl
31	169	16.5	1161	2	US-08-482-293A-53 Sequence 53, Appl
32	169	16.5	1161	2	US-08-943-363-53 Sequence 53, Appl
33	169	16.5	1161	4	US-09-193-043-53 Sequence 53, Appl
34	169	16.5	1161	4	US-09-688-307A-53 Sequence 53, Appl
35	164	16.0	2813	3	US-08-485-618-93 Sequence 93, Appl
36	164	16.0	2813	3	US-09-132-652-2 Sequence 2, Appl
37	161	15.7	264	1	US-08-485-618-93 Sequence 93, Appl
38	161	15.7	264	1	US-08-362-652-93 Sequence 93, Appl
39	161	15.7	264	2	US-08-605-672-93 Sequence 93, Appl
40	161	15.7	264	2	US-08-482-293A-93 Sequence 93, Appl
41	161	15.7	264	2	US-08-943-363-93 Sequence 93, Appl
42	161	15.7	264	4	US-09-193-043-93 Sequence 93, Appl
43	161	15.7	264	4	US-09-688-307A-93 Sequence 93, Appl
44	160	15.6	1161	1	US-08-172-497-2 Sequence 2, Appl
45	160	15.6	1161	1	US-08-286-889-2 Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-462-128-37
; Sequence 37, Application US/08462128
; Patent No. 5686059
; GENERAL INFORMATION:
; APPLICANT: Goetlinck, Paul F.
; APPLICANT: Tondra, David, Mehridad
; TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESS: Lohive & Cockfield
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,128
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/006,096
; FILING DATE: 15-JAN-1993
; APPLICATION NUMBER: US 07/866,403
; FILING DATE: 10-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: MCP-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-462-128-37
;
; Query Match 20.0% Score 205; DB 1; Length 496;
; Best Local Similarity 33.5%; Pred. NO. 4.4e-17;
; Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;
;
; QY 28 DVIIVLDGNSIYR--WSEVQFLRLVGLKFLIDPEIQVGLVQYGSPEVHWSLGDPRRT 85
```

Db 275 DLVFLIDGSKSVRPENFELVKRISQIVTLVDSDLAQVLQVSSVRQEPFLGRFHT 334
QY 86 KEEVVRRAAKNLRRREGRETAKTQAQIMVACTEGFSQSHGCRPEARLLVVTVDGESHGDEE 145
Db 335 KKDIIKAAYRNMSYME-KGTMTGAAALKYLLDINSFTVSSGARPGAKQKGIIVFTDGRSODIYN 393
QY 146 LPAALAKCEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRTIASDP-DERFF 196
Db 394 -DAKAKKDLGFKMFVAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 2
US-08-463-180-37
Sequence 37, Application US/08463180
Patent No. 5741670
GENERAL INFORMATION:
APPLICANT: Goelink, Paul F.
APPLICANT: Tondra, M., Mehrdad
TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
TITLE OF INVENTION: USE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,180
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/006,096
FILING DATE: 15-JAN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/866,403
FILING DATE: 10-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-005DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-463-180-37

Query Match 20.0%; Score 205; DB 1; Length 496;
Best Local Similarity 33.5%; Pred. No. 4.4e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

QY 28 DVIYVLDGNSIYIP-WSEVOTFLRLVGLFIDPEQIQVGLVQGESPVHEMSIGDFRT 85
Db 275 DLVFLIDGSKSVRPENFELVKRISQIVTLVDSDLAQVLQVSSVRQEPFLGRFHT 334
QY 86 KEEVVRRAAKNLRRREGRETAKTQAQIMVACTEGFSQSHGCRPEARLLVVTVDGESHGDEE 145
Db 335 KKDIIKAAYRNMSYME-KGTMTGAAALKYLLDINSFTVSSGARPGAKQKGIIVFTDGRSODIYN 393
QY 146 LPAALAKCEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRTIASDP-DERFF 196

Db 394 -DAKAKKDLGFKMFVAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 3
US-08-001-078A-1
Sequence 1, Application US/08001078A
Patent No. 5872094
GENERAL INFORMATION:
APPLICANT: Goelink, Paul F.
APPLICANT: Tondra, M., Mehrdad
TITLE OF INVENTION: METHODS FOR PROMOTING CARTILAGE MATRIX
TITLE OF INVENTION: FORMATION
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, Suite 510
CITY: BOSTON
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/001,078A
FILING DATE: 06-JAN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-001-078A-1

Query Match 20.0%; Score 205; DB 2; Length 496;
Best Local Similarity 33.5%; Pred. No. 4.4e-17;
Matches 58; Conservative 33; Mismatches 64; Indels 18; Gaps 6;

QY 28 DVIYVLDGNSIYIP-WSEVOTFLRLVGLFIDPEQIQVGLVQGESPVHEMSIGDFRT 85
Db 275 DLVFLIDGSKSVRPENFELVKRISQIVTLVDSDLAQVLQVSSVRQEPFLGRFHT 334
QY 86 KEEVVRRAAKNLRRREGRETAKTQAQIMVACTEGFSQSHGCRPEARLLVVTVDGESHGDEE 145
Db 335 KKDIIKAAYRNMSYME-KGTMTGAAALKYLLDINSFTVSSGARPGAKQKGIIVFTDGRSODIYN 393
QY 146 LPAALAKCEAG-RVTRYGIAVLGHYLRQRDPSSFLEIRTIASDP-DERFF 196
Db 394 -DAKAKKDLGFKMFVAVGVG-----NAVEDELREIASEPVAEHYFY 433

RESULT 4
US-08-897-443-4
Sequence 4, Application US/08897443
Patent No. 5981263
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
APPLICANT: Kaser, Mathew
TITLE OF INVENTION: HUMAN MATRILIN-3


```
QY      28 DNVVLIDGNSLYR--WSEVQTEFLRLAYCKLFIDREQLQUGLGVQYESRPENHESLSDFT 85
Db      275 DLVFILDGSKSVRPENFELVKKFIISQIVLTLDVSDLAOLGLVQYSSVRQERPLRFHT 334
QY      86 KEEVVRAANLSRRREGRTKTQAIIWACTCEGSOSHGRPRPAARLLVUTDDESHDGE 145
Db      335 KDIDIAAVNMMSYME-KGTMTGAALXYLLIDNSTVSGABRGOKGVILEFTDRSODIYN 393
QY      146 LPALAKACEAG-RVTTRYGI AVLGNHYLRORDRPSFLREIRTIASDP-DEEFF 196
Db      334 -DAKKAKDLGKRMFAVG-----NAVDLEIREISEVALEIFY 433
```

```

1      RESULT 7
2      US-08-485-618-101
3      ; Sequence 101: Application US/08485618
4      ; Patent No. 572853
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Gallatin, W. Michael
7      ; APPLICANT: Van der Vieren, Monica
8      ; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
9      ; NUMBER OF SEQUENCES: 103
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
12     ; STREET: 233 South Wacker Drive, 6300 Sear Tower
13     ; CITY: Chicago
14     ; STATE: Illinois
15     ; COUNTRY: United States
16     ; ZIP: 60606-6402
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: Patentln Release #1.0, Version #1.25
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/485,618
24     ; FILING DATE:
25     ; CLASSIFICATION: 435
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 08/173,497
28     ; FILING DATE: 23-DEC-1993
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: US 08/286,889
31     ; FILING DATE: 5-AUG-1994
32     ; PRIOR APPLICATION DATA:
33     ; APPLICATION NUMBER: US 08/362,652
34     ; FILING DATE: 21-DEC-1994
35     ; ATTORNEY/AGENT INFORMATION:
36     ; NAME: Williams Jr., Joseph A.
37     ; REGISTRATION NUMBER: 38,659
38     ; REFERENCE/DOCKET NUMBER: 27866/32797
39     ; TELECOMMUNICATION INFORMATION:
40     ; TELEPHONE: 312-474-6300
41     ; TELEFAX: 312-474-0448
42     ; TELEX: 25-3856
43     ; INFORMATION FOR SEQ ID NO: 101:
44     ; SEQUENCE CHARACTERISTICS:
45     ; LENGTH: 413 amino acids
46     ; TYPE: amino acid
47     ; TOPOLOGY: linear
48     ; MOLECULE TYPE: protein
49     ; US-08-485-618-101

```

[illegible]

Db 177 KD-TNQFSLMQSNVLVTHFTFSSFRSSNPQGLVEPIVQLTGL-TFTATGILIKVTEL 235

Qy 118 FSSSHGREGREARALIVWTGDESH-DGRELPAALAKAGEARVRYGVAIVGHYILRRQDP 176

Db 236 FQYRKARSSAKILIVITDQGYKDPRLHYSAVLPOAEQAGITKVALGVDAF-----QKP 251

Qy 177 SSFRLRIRTIASDPDERFFENV 198

Db 292 TA-ROELDTIASPPDAHVFOV 312

```

1      RESULT 8
2      US-08-605-672-101
3      ; Sequence 101, Application US/08605672
4      ; Patent No. 5817515
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Gallatin, W. Michael
7      ; APPLICANT: Van der Vieren, Michica
8      ; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
9      ; NUMBER OF SEQUENCES: 103
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
12     ; STREET: 233 South Wacker Drive, 6300 Seear Tower
13     ; City: Chicago
14     ; STATE: Illinois
15     ; COUNTRY: United States
16     ; ZIP: 60606-6402
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: IBM floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: Patentin Release #1.0, Version #1.25
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/08/605,672
24     ; FILING DATE:
25     ; CLASSIFICATION: 530
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 08/173,497
28     ; FILING DATE: 23-DEC-1993
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: US 08/286,889
31     ; FILING DATE: 5-AUG-1994
32     ; PRIOR APPLICATION DATA:
33     ; APPLICATION NUMBER: US 08/362,652
34     ; FILING DATE: 21-DEC-1994
35     ; ATTORNEY/AGENT INFORMATION:
36     ; NAME: Williams Jr., Joseph A.
37     ; REGISTRATION NUMBER: 38,659
38     ; REFERENCE/DOCKET NUMBER: 27866/32684
39     ; TELECOMMUNICATION INFORMATION:
40     ; TELEPHONE: 312-474-6300
41     ; TELEFAX: 312-474-0448
42     ; TELEX: 25-3856
43     ; INFORMATION FOR SEQ. ID NO.: 101:
44     ; SEQUENCE CHARACTERISTICS:
45     ; LENGTH: 413 amino acids
46     ; TYPE: amino acid
47     ; TOPOLOGY: linear
48     ; MOLECULE TYPE: protein
49     ; US-08-605-672-101

```

[illegible]

QY 118 FSQSHGGRPEARLLVVTDDGSH-DGELPALAKACEAGVTRYGIAVLGHYLRQDP 176
Db 236 FQTKGARESAKKILIVITDQCKYKDPILYSVAVIPQAEQAGIRYAVIGVDAF----QKP 291
QY 177 SSFLREIRTIASDPERFFENV 198
Db 292 TA-ROELDTIASPPDAHVFQV 312

RESULT 9

US-08-482-293A-101
; Sequence 101, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-293A-101

Query Match 19.3%; Score 198; DB 2; Length 413;
Best Local Similarity 29.7%; Pred. No. 2.5e-16;
Matches 60; Conservative 38; Mismatches 92; Indels 12; Gaps 8;

QY 1 GICARVDASFOPOGSLAFTACRCPTY-MDVIVILDGNSNI--YPMSEVQFLRLVGLKF 57
Db 119 GFCVILDAHAPDPIGVPAALPECPDQEMDIVFLIDSGSISNDPRKMKDFVRAVMDQ-F 177
QY 58 IDPEOIGVLVOYGESPVHEWSLDGFRTEKEVVRRAKNLSRREGRETKYTAQAINVACTEG 117
Db 178 KD-TNTQSLMOYSNVLVTHFTFSSFRNSNPQGLVEPIVQLTGL-TFTAIGILKVTEL 235
QY 118 FSQSHGGRPEARLLVVTDDGSH-DGELPALAKACEAGVTRYGIAVLGHYLRQDP 176

Db 236 FQTKGARESAKKILIVITDQCKYKDPILYSVAVIPQAEQAGIRYAVIGVDAF----QKP 291
QY 177 SSFLREIRTIASDPERFFENV 198
Db 292 TA-ROELDTIASPPDAHVFQV 312

RESULT 10

US-08-943-363-101
; Sequence 101, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,363
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 101:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 413 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-943-363-101

Query Match 19.3%; Score 198; DB 2; Length 413;
Best Local Similarity 29.7%; Pred. No. 2.5e-16;
Matches 60; Conservative 38; Mismatches 92; Indels 12; Gaps 8;

QY 1 GICARVDASFOPOGSLAFTACRCPTY-MDVIVILDGNSNI--YPMSEVQFLRLVGLKF 57
Db 119 GFCVILDAHAPDPIGVPAALPECPDQEMDIVFLIDSGSISNDPRKMKDFVRAVMDQ-F 177
QY 58 IDPEOIGVLVOYGESPVHEWSLDGFRTEKEVVRRAKNLSRREGRETKYTAQAINVACTEG 117
Db 178 KD-TNTQSLMOYSNVLVTHFTFSSFRNSNPQGLVEPIVQLTGL-TFTAIGILKVTEL 235
QY 118 FSQSHGGRPEARLLVVTDDGSH-DGELPALAKACEAGVTRYGIAVLGHYLRQDP 176
Db 236 FQTKGARESAKKILIVITDQCKYKDPILYSVAVIPQAEQAGIRYAVIGVDAF----QKP 291

STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 2072792
US-08-897-443-3

Query Match 19.1%; Score 196; DB 2; Length 956;
Best Local Similarity 30.2%; Pred. No. 1.7e-15;
Matches 64; Conservative 33; Mismatches 83; Indels 32; Gaps 8;

QY 6 VDASQPOGSLPTAORCTY-----MDVTVLDGNSI--YPMSEVQFTL 49
DB 21 VGRERPOARFPRSGRHVMYDOTALLESSCENKPADLVFIIDSSRSVNTYDAKKEFI 80
QY 50 RLAVKLFIDPQIOVLQVQGESPVHEMSIGDFRTKEVRAAKNLSRREGRETKTQA 109
DB 81 LDILQFLDIDPVTYVGLQVSTVKNESLKTFRKSEVERAVKRM-RHLSTGTMTGLA 139
QY 110 IMVACTEGFSQSHGRP---EAARLLVVTDGSHDEELPALAKACEAGVTRYGIATL 166
DB 140 IQYALNIAFSEAEAGAPLEENVPRIMITDGRPDQSAVEAA-KARNTG-ILIFALGV- 196
QY 167 GHYLRQRPDSSFLREIRTIADPDPERFFNV 198
DB 197 -----QVVD---LNTLKAIGSEPHKHVFLV 219

RESULT 14

US-08-897-443-1
Sequence 1, Application US/08897443
Patent No. 5981263

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Lal, Preeti

APPLICANT: Corley, Neil C.

APPLICANT: Shah, Purvi

APPLICANT: Kaser, Mathew

TITLE OF INVENTION: HUMAN MATRILIN-3

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Inocyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/897,443

FILING DATE: Filed Herewith

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0348 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 638 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: UTRSN0702

CLONE: 681719
US-08-897-443-1

Query Match 18.6%; Score 191; DB 2; Length 638;
Best Local Similarity 33.0%; Pred. No. 3.8e-15;
Matches 58; Conservative 33; Mismatches 67; Indels 18; Gaps 7;

QY 28 DVIYVLDGNSI--YPMSEVQFTLRVLGKFLIDPEQIQVLQVQGESPVHEMSIGDFRT 85
DB 57 DLVFIIDSSRSRVNTHDAKKEFLVDILQFLDIDPVTYVGLQVSTVKNESLKTFRK 116
QY 86 KEEVRAAKNLSRREGRETKTQAQIMVACTEGFSQSHGRP---EAARLLVVTDGSHD 142
DB 117 KSEVERAVKRM-RHLSTGTMTGLAIQYALNIAFSEAEAGAPLEENVPRIMIVTDGRPD 175
QY 143 GEEPLPAKACEAGVTRYGIATVGHYLRQRPDSSFLREIRTIADPDPERFFNV 198
DB 176 SVAEVA-KARDTG-ILIFALGV-----QVVD---FNTLKSIGSEPHKHVFLV 219

RESULT 15

US-08-199-776-2
Sequence 2, Application US/08199776

Patent No. 5594120

GENERAL INFORMATION:

APPLICANT: Brenner, Michael B.

APPLICANT: Parker, Christina M.

TITLE OF INVENTION: No. 5594120e1 integrin alpha subunit

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield and Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/199,776

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Plumer, Elizabeth R.

REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7020

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1178 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-199-776-2

Query Match 17.3%; Score 177.5; DB 1; Length 1178;
Best Local Similarity 31.0%; Pred. No. 5.1e-13;
Matches 54; Conservative 31; Mismatches 80; Indels 9; Gaps 5;

QY 28 DVIYVLDGNSIYP--WSEVQFTLRVLGKFLIDPEQIQVLQVQGESPVHEMSIGDFRT 85
DB 201 EIALIILDSGSDIDPDQFQARDFISNMNRNYEKCFCFENFALVOYGIVQTEFLRDSOD 260
QY 86 KEEVRAAKNLSRREGRETKTQAQIMVACTEGFSQSHGRPREARLLVVVTDGE-SHGE 144
DB 261 VMASIAVQNTT-QVGSVTKTASAMQHVLDISFTSSHSRKAQKAWVVLTDGGIFEDPL 319
QY 145 ELPAALKACEAGVTRYGIATVGHYLRQRPDSSFLREIRTIADPDPERFFNV 198

Db 320 NLTVINSPKMOGVERFAIGV-GEFFKSART---ARELNLIASDPDETHAFKV 368

Search completed: July 16, 2003, 07:57:47
Job time : 8.79747 secs

C; Superfamily: integrin beta chain; laminin-type EGF-like homology

Db 17 PAC--AAVARLRLGETVHAFAHFDV-----EVIGAIRQAAVRQLISDHEGL 60
QY 65 VGIYVQGESFVHMSLGDFTKEEVVRAAKNLSRREGRETKTAQAL---MVACTEGFSQS 121
Db 61 VVVVNVFLSLPVRWRPLKPFQRAYQURSTHTVA--DGAVVALAELGLVPLITCDGLAOS 118
QY 122 HGGPPE 127
Db 119 HGHNAE 124

RESULT 11
G95210
acetyltransferase, GNAT family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: G95210
R:Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95210
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KUR>
A:Cross-references: GB:AE005672, PIDN:AAK75880.1, PID:G14973106, GSPDB:GN00164, TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1807

Query Match 6.7%; Score 68.5; DB 2; Length 148;
Best Local Similarity 25.0%; Pred. No. 45;
Matches 28; Conservative 25; Mismatches 34; Indels 25; Gaps 6;

QY 9 SFQPGSLAPTAQRCPYMDVIVLDGNSIYPMSEVQTFRLRL-----VGKLFIDP 60
Db 27 SFQEQW-MGP---RIP-FLITLQALLEGVFSIFDEQFVGFIQIKRLSDSNLHGRFIP 81

QY 61 EQLQVGLVQGESFVHMSLGDFTKE-----EVVRAAKNLSRREGRE 103
Db 82 OKQEQGL--GSKALRKRFVSLAFENDIDISLNVFANQRAQNLVQKGFPE 130

RESULT 12
H75017
hypothetical protein PAB1243 - Pyrococcus abyssi (strain Oresay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: H75017
R:anonymous, Genoscope
Submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: H75017
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-171 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CA850582.1; PID:el51648
A:Experimental source: strain Oresay
C:Genetics:
A:Gene: PAB1243

Query Match 6.6%; Score 68; DB 2; Length 171;
Best Local Similarity 24.3%; Pred. No. 60;
Matches 33; Conservative 17; Mismatches 44; Indels 42; Gaps 6;

QY 40 YKMSVQOTFLRLVGLTFLDPEQIYVGLVQGESP-----VHMSLGDG---- 83
Db 52 YKISHISSAMKYLEG-----VGLVQRIKKPGDRRAVFAVATKNSFSEWSSAFYEKI 101

QY 84 -----RTKEEVVRAAKNLSRREGRETKTAQALMVACTEGFSQSHGGRPEAR-LIIVVTD 137
Db 102 LRDIQETKESIRALAELEEGEGSEVEIEIKETLKAALR-----RNEVARKLLTLIMQ 153
QY 138 GESHDGEELPAALKAC 153
Db 154 FKSE--EELIKVLESC 167

RESULT 13
AF2520
hypothetical protein alr7342 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AF2520
R:Kaneako, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaamoto, S.; Watanabe, A.; Iritguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AF2520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA877100.1; PID:G17134541; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7342
A:Genome: plasmid

Query Match 6.6%; Score 67.5; DB 2; Length 151;
Best Local Similarity 26.9%; Pred. No. 58;
Matches 14; Conservative 19; Mismatches 18; Indels 1; Gaps 1;

QY 50 RRLVGLKFLIDPEQIYVGLVQGESPVH-EWSLGDFTKEEVVRAAKNLSRRE 100
Db 95 KRLVAVEVFNRRSVVLTVVQGGAVVRYQLKGYNTKEQFLQAEANAKQK 146

RESULT 14
E75544
Jag-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: E75544
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.V.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Usterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-193 <WHI>
A:Cross-references: GB:AF001866; GB:AE000513; NID:G6457921; PIDN:AAF09828.1; PID:G645791
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0246
A:Map position: 1

Query Match 6.6%; Score 67.5; DB 2; Length 193;
Best Local Similarity 24.5%; Pred. No. 77;
Matches 40; Conservative 22; Mismatches 50; Indels 51; Gaps 9;

QY 13 QGSLAPTAQRCPYMDVIVLDGNSIYPMSEVQTFRLRLVGLKFLIDPE-QIYVGLVQYG 71
Db 21 ESALPPPA---PDAPVSAPIHGADG-DRAVLQGLFLAELNAR--IDPGLRVQV----- 68
QY 72 ESPVHMSLGDFTKEEVVRA--AKNLSRREGRETKTAQALMVACTEGFSQSHGGRPE- 127

Db 69 -----RETEDALEAELSGENNAARLAGRDRGTGAIEVIAVAVLAK-HAGRGDL 115
 QY 128 -----AARLLV-VYTDGSHDGEELPAA 149
 Db 116 RYRVDPVGFGRKROADTLTKLAERLANVQVAKSGEPHELOPMPAA 158

RESULT 15

H81150
 hypobacterial protein NMB0861 [imported] - Neisseria meningitidis (strain MC58 serogroup C)
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Feb-2001
 C:Accession: H81150; F81872
 R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, S.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755; PMID:10710307
 A:Accession: H81150
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-181 <TET>
 A:Cross-References: GB:AEO02439; GB:AEO02098; NID:g7226100; PIDN:AAFA1272.1; PID:g722609
 A:Experimental source: serogroup B, strain MC58
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 N:Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: A81775; MUID:2022556; PMID:10761919
 A:Accession: F81872
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-181 <PAR>
 A:Cross-References: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84336.1; PID:g737976
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMB0861; NMA1073

Query Match 6.5%; Score 67; DB 2; Length 181;
 Best Local Similarity 23.5%; Pred. No. 80;
 Matches 28; Conservative 17; Mismatches 48; Indels 26; Gaps 3;
 QY 83 FTKEEVVAAKNTLSRREGRETKA-----QAIWVACTGEGFSGSHGRPEARLLV 133
 Db 33 FHTRADAPWQALAELOKEMKETEGAFPLAIIIGAAIGMTQHGFSYATTGRPASVRYVA 92
 QY 134 V-----VTGESHGDEELPAA-----LRACEAGRVTRYGIAVLGHYLRORD 175
 Db 93 IAGGIGAIIGGVGAAGKVVSPFAKYGREIKIGNNMRIAPFGNRTGHPICKKFPYHRRVTD 151

Search completed: July 16, 2003, 08:45:47
 Job time : 42 secs

THIS PAGE BLANK (USPTO)

QY 85 TKEEVRAAKNLSRREGRETQAIVMVACTEGFSQSHGCRPEARLLVVTGDSHDE 144
 Db 57 VNSPEARRELVAMAGNRPKTAAPLVALLSADNFBELPBLF-----P 102
 QY 145 ELPAALACEAGRTVRYGIAVLGHYLR 171
 Db 103 HEPAAKDAFFESRPVREGAATLNAALQ 129

RESULT 2

GON2_CLAGA STANDARD; PRT; 86 AA.
 ID GON2_CLAGA
 AC P43306;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
 DE (GnRH-II) (LH-RH II) (Lutaliberin II).
 GN GNRH2.
 OS Clarias gariepinus (Sharptooth catfish) (African catfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
 OC Clariidae; Clarias.
 OC NCBI_TaxID=13013;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94291651; PubMed=8020492;
 RA Bogerd J., Zandbergen T., Andersson E., Goos H.,
 RT "Isolation, characterization and expression of cDNA encoding the
 catfish-type and chicken-II-type gonadotropin-releasing-hormone
 RT precursors in the African catfish".
 RL Eur. J. Biochem. 222:541-549(1994).
 RN [2]
 RP SEQUENCE OF 25-34.
 RC TISSUE=Brain;
 RX MEDLINE=92392313; PubMed=1520292;
 RA Bogerd J., Li K.W., Janssen-Dommerholt C., Goos H.,
 RT "Two gonadotropin-releasing hormones from African catfish (Clarias
 RT gariepinus)".
 RL Biochem. Biophys. Res. Commun. 187:127-134(1992).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X78047; CAA54969.1; -
 DR PIR: JCI243; RHID2S.
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 86 PROGONADOLIBERIN II.
 FT PEPTIDE 25 34 GONADOLIBERIN II.
 FT PEPTIDE 38 86 GNRH-ASSOCIATED PEPTIDE II.
 FT MOD_RES 25 25 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 34 34 AMIDATION (G-35 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 86 AA; 9766 MW; 4AD9F24597E77EBF CRC64;
 Query Match 6.3%; Score 65; DB 1; Length 86;
 Best Local Similarity 26.8%; Pred. No. 24;
 Matches 22; Conservative 10; Mismatches 20; Indels 30; Gaps 4;
 QY 109 AINVACTE---GFSQ--SHGCRPEARLLVVTGDSHDEELPALKACEAGRTVRYGI 163

Db 11 AALLLCQAQLSFSGHSHGWPGGKREI-----DSYSSPEISGEIKLCEAGE----- 58
 QY 164 AVLGHYLRQDRPSSFLREIRT 185
 Db 59 -----CSYLRPLRT 67

RESULT 3

GON2_HAPBU STANDARD; PRT; 85 AA.
 ID GON2_HAPBU
 AC P37044; P20408;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
 DE (GnRH-II) (LH-RH II) (Lutaliberin II).
 GN GNRH2.
 OS Haplochromis burtoni (Burton's mouthbrooder).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorphi; Perciformes; Labroidae;
 OC Cichlidae; Ascotilapia.
 OC NCBI_TaxID=8153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94151343; PubMed=8108425;
 RA White S.A., Bond C.T., Francis R.C., Kaeten T.L., Fernald R.D.,
 RA Adelman J.P.,
 RT "A second gene for gonadotropin-releasing hormone: cDNA and
 RT expression pattern in the brain".
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1423-1427(1994).
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ONLY ONE CELL GROUP IN THE
 CC MESENCEPHALON.
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L27435; AAA74993.1; -
 DR InterPro: IPR002012; GNRH.
 DR Pfam: PF00446; GNRH; 1.
 DR PROSITE: PS00473; GNRH; 1.
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
 KW Signal; Multigene family.
 FT SIGNAL 1 23
 FT CHAIN 24 85 PROGONADOLIBERIN II.
 FT PEPTIDE 24 33 GONADOLIBERIN II.
 FT PEPTIDE 37 64 GNRH-ASSOCIATED PEPTIDE II-1 (POTENTIAL).
 FT PEPTIDE 67 85 GNRH-ASSOCIATED PEPTIDE II-2 (POTENTIAL).
 FT MOD_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 85 AA; 9631 MW; CF8C0EDBF27365F CRC64;
 Query Match 6.2%; Score 63.5; DB 1; Length 85;
 Best Local Similarity 32.9%; Pred. No. 32;
 Matches 25; Conservative 6; Mismatches 26; Indels 19; Gaps 4;
 QY 110 INVACTEFSQ--SHGCRPEARLLVVTGDSHDEELPALKACEAGRTVRYGIAVLG 167
 Db 14 LCVGAQSLFAQHWGSHGWPGGKREL-----DSFGTSSISEIKLCEAGESS----- 59
 QY 168 HYLRQDRPSSFLREI 183
 Db 60 -YLRPQ--RSILRNI 72

RESULT 4
HIS3_RHIME STANDARD; PRT; 150 AA.
AC Q920B5; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosyl-AMP cyclodiphosphate (EC 3.5.4.19) (PRA-CH).
GN Hist OR R01419 OR SMC01004.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX NCBI_Taxid=382;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gonzy J., Bothe G., Ampe F., Batut J.,
RA Bolteard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandemol M., Weidner S., Gallibert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- PHOSPHORIBOSYL-5-(5-phosphoribosyl)-AMP + H(2)O = 1-(5-phosphoribosyl)-5-(5-phosphoribosylamino)methylideneaminoimidazole-4-carboxamide.
CC -1- PATHWAY: Histidine biosynthesis; third step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PRA-CH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AL591787; CAC45998.1; -
CC DR InterPro; IPR002496; PRA-CH.
CC DR Pfam; PF01502; PRA-CH; 1.
CC DR ProDom; PD002610; PRA-CH; 1.
CC KW Histidine biosynthesis; Hydrolase; Complete proteome.
CC SQ SEQUENCE 150 AA; 16560 MW; 301D4FE9B6D5E83 CRC64;
Query Match 6.2%; Score 63.5; DB 1; Length 150;
Best Local Similarity 35.1%; Pred. No. 63;
Matches 26; Conservative 7; Mismatches 28; Indels 13; Gaps 4;
QY 127 EAARLLVVTGDSHDEELPALAKACEAGRT-RYGAIVLCHYLRORD-----PSS 178
DB 27 EKGITAVYTD--ARDGELVVAHNMNAALSTLTERGA--HYYSRSRDLRMKGESG 81
QY 179 FAREIRITASDPE 192
DB 82 NLQTVREIRTDQDQ 95
RESULT 5
GLBC_CHITH STANDARD; PRT; 161 AA.
AC P12548; P02225;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin CTT-VIIB-3 precursor.
GN CTT-7B3.
OS Chironomus thummi thummi (Midge).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Chironomidae; Chironomidae; Chironominae; Chironomus.
OX NCBI_Taxid=7155;
RN [1]
RS SEQUENCE FROM N.A.
RC MEDLINE=86103336; PubMed=3841283;
RA Saffariani D.A., Trewalt P.M., Caetano M., Wejkenora P.J., Bergtrom G.;
RT "Deoxynucleotide sequence of an insect cDNA codes for an unreported
RT member of the Chironomus thummi globin family".
RL Biochem. Biophys. Res. Commun. 133:641-647(1985).
RN [2]
RS SEQUENCE OF 17-161 (MIXTURE OF ISOZYMES).
RX MEDLINE=79129112; PubMed=422121;
RA Sladic-Simic D., Kleinschmidt T., Brauntz G.;
RT "Hemoglobins, XXVI. Analysis of the primary structure of the dimeric
RT insect haemoglobin CTT VIIB (Erythrocytorin) from Chironomus thummi
RT thummi. Diptera.";
RL Hoppe-Seyler's Z. Physiol. Chem. 360:115-124(1979).
CC -1- SUBUNIT: Homodimer.
CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
CC GLOBIN.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M11620; AAA8260.1; -
CC DR PIR; A02547; GLOBE8.
CC DR HSSP; P02229; ILECA.
CC DR InterPro; IPR002336; Erythrocytorin.
CC DR InterPro; IPR000971; Globin.
CC DR Pfam; PF00042; globin; 1.
CC DR PRINTS; PR00611; ERYTHROCYTORIN.
CC DR PROSITE; PS01033; GLOBIN; 1.
CC KW Heme; Oxygen transport; Transport; Multigene family;
CC KW Signal.
CC FT SIGNAL 1 16
CC FT CHAIN 17 161 GLOBIN CTT-VIIB-3.
CC FT METAL 76 76 IRON (HEME DISTAL LIGAND)
CC FT METAL (BY SIMILARITY).
CC FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)
CC FT (BY SIMILARITY).
CC SQ SEQUENCE 161 AA; 16983 MW; 7A3698ADF138DD93 CRC64;
Query Match 6.0%; Score 62; DB 1; Length 161;
Best Local Similarity 29.5%; Pred. No. 95;
Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2;
QY 8 ASFPOGSLAPTAQRCPTVMVIVLDSNSIYPSVQTFRLRLVGLFIDPEIQVGL 67
DB 65 ASIKDTGAFARHARIVFSLSEVALASGNES-----NASVNSLVSLGPDHAKRGVSA 118
QY 68 VOYGESPVHEWSLGDFT 85
DB 119 AQFGE-----FRT 126
RESULT 6
GLBF_CHITH STANDARD; PRT; 161 AA.
AC P12549; P02225;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin CTT-VIIB-6 precursor.
GN CTT-7B6.
OS Chironomus thummi thummi (Midge).
OC

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.
 OK NCBI_TaxID=7155;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89137998; PubMed=2852146;
 RA Trewitt P.M., Saffarini D.A., Bergtrom G.;
 RT "Multiple clustered genes of the haemoglobin VIIB subfamily of
 RT Chironomus thummi thummi (Diptera).";
 RL Gene 69:91-100(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96054034; PubMed=7563117;
 RA Trewitt P.M., Luhm R.A., Samad F., Ramakrishnan S., Kao W.Y.,
 RA Bergtrom G.;
 RT "Molecular evolutionary analysis of the YWVZ/7B globin gene cluster
 RT of the insect Chironomus thummi.";
 RL J. Mol. Evol. 41:313-328(1995).
 RN [3]
 RP SEQUENCE OF 17-161 (MIXTURE OF ISOZYMES).
 RX MEDLINE=79129112; PubMed=422121;
 RA Sladic-Simic D., Kleinschmidt T., Braunitzer G.;
 RT "Hemoglobins, XXVI. Analysis of the primary structure of the dimeric
 RT insect haemoglobin CTT VIIb (Erythrocytorin) from Chironomus thummi
 RT thummi, Diptera.";
 RL Hoppe-Seiler's Z. Physiol. Chem. 360:115-124(1979).
 CC -1- SUBUNIT: Homodimer.
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
 CC GLOBIN.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U07703; AAA85486.1; -
 DR PIR: A02547; G6ICE8.
 DR PIR: A30477; A30477.
 DR HSSP: P02229; IECA.
 DR InterPro: IPR002336; Erythrcurin.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin.1.
 DR PRINTS: PR00611; ERYTHRCURIN.
 DR PROSITE: PS01033; GLOBIN.1.
 KW Heme; Oxygen transport; Transport; Multigene family;
 KM Signal.
 FT SIGNAL. 1 16
 FT CHAIN 17 161 GLOBIN CTT-VIIB-6.
 FT METAL 76 76 IRON (HEME DISTAL LIGAND)
 FT METAL 76 76 (BY SIMILARITY).
 FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)
 FT METAL 111 111 (BY SIMILARITY).
 SQ SEQUENCE 161 AA; 16959 MW; EAIBBC54BA432EF CRC64;
 Query Match 6.0%; Score 62; DB 1; Length 161;
 Best Local Similarity 29.5%; Pred. No. 95;
 Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2;
 QY 8 ASFGQGSGLAPTAQRCPTMDVIVLDGNSIYPMSEVOTFLRLVYGLFIDPEIQVGL 67
 DB 65 ASIKDTGAFATHTATVTSFLSEVIALSGNDS-----NAAVNSLVSKLGDHKGAGVSA 118
 OY 68 VOYGESPVHWSLGDERT 85
 DB 119 AQFG-----FRT 126

RESULT 7
 GBLI CHITP
 ID GBLI CHITP STANDARD; PRT; 161 AA.
 AC 023763;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Globin CTT-VIIB-8 precursor.
 GN CTT-7B8.
 OS Chironomus thummi thummi (Midge).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.
 OK NCBI_TaxID=7155;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96054034; PubMed=7563117;
 RA Trewitt P.M., Luhm R.A., Samad F., Ramakrishnan S., Kao W.Y.,
 RA Bergtrom G.;
 RT "Molecular evolutionary analysis of the YWVZ/7B globin gene cluster
 RT of the insect Chironomus thummi.";
 RL J. Mol. Evol. 41:313-328(1995).
 CC -1- SUBUNIT: Homodimer.
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
 CC GLOBIN.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U07703; AAA85490.1; -
 DR HSSP: P02229; IECA.
 DR InterPro: IPR002336; Erythrcurin.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin.1.
 DR PRINTS: PR00611; ERYTHRCURIN.
 DR PROSITE: PS01033; GLOBIN.1.
 KW Heme; Oxygen transport; Transport; Multigene family;
 KM Signal.
 FT SIGNAL. 1 16
 FT CHAIN 17 161 GLOBIN CTT-VIIB-8.
 FT METAL 76 76 IRON (HEME DISTAL LIGAND)
 FT METAL 76 76 (BY SIMILARITY).
 FT METAL 111 111 IRON (HEME PROXIMAL LIGAND)
 FT METAL 111 111 (BY SIMILARITY).
 SQ SEQUENCE 161 AA; 17040 MW; C78B9E70C1F7E161 CRC64;
 Query Match 6.0%; Score 62; DB 1; Length 161;
 Best Local Similarity 29.5%; Pred. No. 95;
 Matches 23; Conservative 10; Mismatches 29; Indels 16; Gaps 2;
 QY 8 ASFGQGSGLAPTAQRCPTMDVIVLDGNSIYPMSEVOTFLRLVYGLFIDPEIQVGL 67
 DB 65 ASIKDTGAFATHTATVTSFLSEVIALSGNES-----NASAVNSLVSKLGDHKGAGVSA 118
 OY 68 VOYGESPVHWSLGDERT 85
 DB 119 AQFG-----FRT 126
 RESULT 8
 GBLI CHITP
 ID GBLI CHITP STANDARD; PRT; 161 AA.
 AC P18966;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)


```

RESULT 12
MT2B_LYCES STANDARD; PRT; 82 AA.
ID MT2B_LYCES
AC 040158; 043514;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Metallothionein-like protein type 2 B.
GN MTB.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxId=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Ailsa Craig;
RA Whiteleaf C.A.; Lelhuquet J.A.; Thurman D.A.; Tomsett A.B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Bonner Beste; TISSUE=Root;
RA Girtich A.; Herdik A.; Balzer H.; Stephan U.; Baumlein H.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: METALLOTHIONEIN HAVE A HIGH CONTENT OF CYSTEINE
CC RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -1- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77966; AAB04675.1; -
DR EMBL; Z68138; CAA92243.1; -
DR InterPro; IPR000347; Metallothion_15.
DR Pfam; PF01439; Metallothio_2; 1.
DR ProDom; PD001611; Metallothion_15; 1.
DR Metal-binding; Metal-thiolate cluster; Multigene family.
FT CONFLICT 17 17 G -> D (IN REF. 2).
SQ SEQUENCE 82 AA; 8253 MM; 68298C4915CE495F CRC64;

Query Match 5.9%; Score 60.5; DB 1; Length 82;
Best Local Similarity 37.0%; Pred. No. 57;
Matches 17; Conservative 7; Mismatches 17; Indels 5; Gaps 2;

QY 35 GSNSTIYP---MSGVOTFLRLVGLFIDPEOIGVGVGSGSPVHE 77
Db 21 GGCNYPDMSTYESTTETTV--LGVPEKTSFGAMGESPVAE 64

RESULT 13
GN2_DICLA STANDARD; PRT; 85 AA.
ID GN2_DICLA
AC Q91A08;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Gonadoliberin II precursor (Gonadotropin-releasing hormone II)
DE (GNRH-II) (LH-RH II) (Laliberin II).
GN GNRH2.
OS Dicotylarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes; Percoidae;
OC Moronidae; Dicentrarchus.
OX NCBI_TaxId=13489;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Brain;
RX MEDLINE=20540016; PubMed=11086295;
RA Gonzalez-Martinez D.; Madigou T.; Zmora N.; Anglade I.; Zanny S.;
RA Zohar Y.; Elizur A.; Munoz-Cueto J.A.; Kah O.;
RT "Differential expression of three different prepro-GNRH
RT (gonadotropin-releasing hormone) messengers in the brain of the
RT European sea bass (Dicentrarchus labrax).";
RL J. Comp. Neurol. 429:144-155(2001).
CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF224281; AAF62900.1; -
DR InterPro; IPR002012; GNRH.
DR Pfam; PF00446; GNRH; 1.
DR PROSITE; PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal.
FT SIGNAL 1 23
FT CHAIN 24 85 BY SIMILARITY.
FT PEPTIDE 24 33 GONADOLIBERIN II.
FT PEPTIDE 37 85 GNRH-ASSOCIATED PEPTIDE II.
FT MOD_RES 24 24 PYROLIDONE CARBOXYLIC ACID
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
FT MOD_RES 33 33 SIMILARITY).
SQ SEQUENCE 85 AA; 9646 MM; F832C0698C842C64 CRC64;

Query Match 5.9%; Score 60.5; DB 1; Length 85;
Best Local Similarity 34.9%; Pred. No. 59;
Matches 22; Conservative 4; Mismatches 20; Indels 17; Gaps 3;

QY 121 SHGPEPARLLVVTGDSHGDELPAALFACEGRTYGVIAVLGHYLRQDPSSFL 180
Db 27 SHGWPGGKREL-----DSFGTSEISEIKLCEGEGCS-----YLPQR--RSLV 69

QY 181 REI 183
Db 70 RNI 72

RESULT 14
UTER_RAT STANDARD; PRT; 96 AA.
ID UTER_RAT
AC P17559;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Clara cell phospholipid-binding protein precursor (CCBPp) (Clara cells
DE 10 kDa secretory protein) (CC10) (Uteroglobin) (PCB-binding protein).
GN SCGB1A1 OR UGB OR CC10 OR UTG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Katyal S.L.; Singh G.; Brown W.E.; Kennedy A.L.; Squeglia N.;
RA Wong-Chong M.-L.;
RT "Clara cell secretory (10 kDa) protein: amino acid and cDNA
RT nucleotide sequences and developmental expression.";
RL Prog. Respir. Res. 25:29-35(1990).
RN [2]
RP SEQUENCE FROM N.A.

```

RA	MedLine=90324266; PubMed=2115524;
RA	Nordlund-Moeller L., Andersson O., Ahlgren R., Schilling J.,
RT	Gillner M., Gustafsson J.-A., Lund U.,
RT	"Cloning, structure, and expression of a rat binding protein for
RT	polychlorinated biphenyls. Homology to the hormonally regulated
RT	progesterone-binding protein uteroglobin.";
RT	J. Biol. Chem. 265:12690-12693(1990).
RM	[3]
RM	SEQUENCE OF 1-18 FROM N.A.
RP	MedLine=90272398; PubMed=2349092;
RX	Hagen G., Wolf M., Katyal S.L., Singh G., Beato M., Suske G.;
RA	"Tissue-specific expression, hormonal regulation and 5'-flanking gene
RT	region of the rat Clara cell 10 kDa protein: comparison to rabbit
RT	uteroglobin.";
RL	Nucleic Acids Res. 18:2939-2946(1990).
RM	[4]
RP	X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX	MedLine=92219263; PubMed=1560460;
RA	Umland T.C., Swannathan S., Furey W., Singh G., Fletcher J., Sax M.;
RT	"Refined structure of rat Clara cell 17 kDa protein at 3.0-A
RT	resolution.";
RT	J. Mol. Biol. 224:441-448(1992).
RM	[5]
RP	STRUCTURE BY NMR.
RX	MedLine=96069785; PubMed=7583672;
RA	Haerd T., Barnes H.U., Larsson C., Gustafsson J.-A., Lund U.;
RT	"Solution structure of a mammalian PCB-binding protein in complex
RT	with a PCB.";
RT	Nat. Struct. Biol. 2:983-989(1995).
CC	-1- FUNCTION: BINDS PHOSPHATIDYLCOLINE, PHOSPHATIDYLINOSITOL,
CC	POLYCHLORINATED BIPHENYLS (PCB) AND WEAKLY PROGESTERONE, POTENT
CC	INHIBITOR OF PHOSPHOLIPASE A2.
CC	-1- SUBUNIT: Homodimer; antiparallel disulfide-linked.
CC	-1- TISSUE SPECIFICITY: CLARA CELLS (NONCILIATED CELLS OF THE
CC	SURFACE EPITHELIUM OF THE PULMONARY AIRWAYS).
CC	-1- INDUCTION: BY GLUCOCORTICOIDS.
CC	-1- SIMILARITY: BELONGS TO THE UTEROGLIBIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on items
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; J05536; AAA41817.1; -;
DR	EMBL; X51318; CA835701.1; -;
DR	PIR; A36581; A36581.
DR	PIR; S10185; S10185.
DR	PIR; S21676; S21676.
DR	PDB; 1CCD; 31-JAN-94.
DR	PDB; 1UTR; 07-DEC-95.
DR	InterPro: IPR003628; Uteroglbn sub.
DR	InterPro: IPR003329; uteroglobin.
DR	Pfam; PF01099; Uteroglobin.1.
DR	PRINTS; PR00486; UTEROGLIBIN.
DR	ProDom; PD012475; Uteroglbn_sub; 1.
DR	SMART; SMO0036; UTG; 1.
DR	PROSITE; PS00403; UTEROGLIBIN_1; 1.
DR	PROSITE; PS00404; UTEROGLIBIN_2; 1.
KW	Phospholipase A2 inhibitor; Signal; 3D-structure.
FT	SIGNAL 1 19
FT	CHAIN 20 96
FT	DISULFID 24 24 CLARA CELL PHOSPHOLIPID-BINDING PROTEIN.
FT	DISULFID 90 90 INTERCHAIN (WITH C-90').
FT	TURN 21 22 INTERCHAIN (WITH C-24').
FT	HELIX 25 34
FT	TURN 35 36
FT	HELIX 39 46
FT	TURN 47 49
FT	HELIX 53 68
FT	HELIX 71 84

FT	TURN	85	86
FT	TURN	88	90
SO	SEQUENCE	96 AA; 10449 MW;	1A12988677B9EBEF CRC64;
Query Match			
Best Local Similarity		5.9%; Score 60.5;	DB 1; Length 96;
Matches		32.4%; Pred. No. 69;	
		Conservative 14;	Mismatches 17; Indels 19; Gaps 5
QY	23	CPTAMDV-IVLDGNSIY-----PMSEVC--FTRRLVGLKFLDPDQIQGVLOY	70
		: : : : : : : : : : : : : : : : : : :	
Db	24	CPGELQVIEALLLTSSENYSNEALKPENPASPLOAGTQLKRLLVDTL--PETRINIVKL	80
		: : : : : : : : : : : : : : : : : : :	
QY	71	GE----SPVHEWSL	80
		: : : : : : : : : : : : : : : : : :	
Db	81	TEKITLSPLCEODL	94

RESULT 15	V250_FOWPV	STANDARD;	PRT; 140 AA.
ID	_V250_FOWPV		
AC	P14362;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Hypothetical protein FV250 (BamHI-ORF4).		
GN	FPV250.		
OS	Fowlpox virus (FPV).		
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;		
OC	Avipoxvirus.		
OX	NCBI_TaxId=10261;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FP-9 / Isolate HP-438;		
RX	MEDLINE=86229622; PubMed=2836548;		
RA	Tomley F., Binn M., Campbell J., Boursnell M.E.G.;		
RT	"Sequence analysis of an 11.2 kilobase, near-terminal, BamHI fragment		
RT	of fowlpox virus.";		
RL	J. Gen. Virol. 69:1025-1040(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20193820; PubMed=1072915;		
RA	Afonso C.L., Tulman E.R., Lu Z., Zsak L., Kutish G.F., Rock D.L.;		
RT	"The genome of fowlpox virus.";		
RL	J. Virol. 74:3815-3831(2000).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; D00295; BAA00196.1; -.		
DR	EMBL; AF198100; AAF44594.1; -.		
DR	PIR; D29963; MWVZF4.		
KW	Hypothetical protein; Early protein		
SO	SEQUENCE 140 AA; 16507 MW; 09094006BC510C425 CXC64;		

Query Match	5.9%	Score 60.5	DB 1	length 140
Best Local Similarity	22.7%	Pred. No. 1.1e+02		
Matches	22	Conservative	20	Mismatches 28; Indels 27; Gaps 5
QY	24	PTWMDVIVLDGSSISIVPWSVQTFILRLVYKFLI-DPEOI-----	QVGLVQYG	71
DB	27	PTRRDIIIGWMDSD--ISMENLIIIMSDSKITIVYDDEALYKXADTMEBSSEIGLIMG		84
QY	72	E-----SPVHWSLGDFTRYEEVYRAAKNLSSR		99
DB	85	NEVYHCREDIKLPE-----EDRCDDEYIMKIREKARQ		117

Wed Jul 16 08:48:45 2003

us-09-647-544-2_copy_140_337.closed.rsp

Page 9

Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:40:06 ; Search time 79 Seconds

(without alignments)
516.422 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025
Sequence: 1 GICARVDASFQPGSLAPTA.....FLREIRIASDPDERFFENV 198

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 284930

Minimum DB seq length: 0

Maximum DB seq length: 198

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	336	32.8	191	6 Q29124	Q29124 sue scrofa
2	92	9.0	146	5 O96930	O96930 plasmidium
3	74	7.2	194	13 Q91415	Q91415 gallus gall
4	72.5	7.1	134	11 Q920X7	Q920X7 cavia porce
5	71.5	7.0	108	17 O8TW99	O8TW99 methanopyru
6	71	6.9	164	16 O9RS46	O9RS46 deinococcus
7	71	6.9	169	17 O58027	O58027 pyrococcus
8	71	6.9	170	6 Q28796	Q28796 pan troglod
9	71	6.9	174	16 Q9X957	Q9X957 streptomyce
10	71	6.9	197	10 Q9M0V0	Q9M0V0 arabidopsis
11	70.5	6.9	118	16 O8ZME3	O8ZME3 salmonella
12	70.5	6.8	118	16 O8Z448	O8Z448 salmonella
13	70	6.9	188	16 O8YH81	O8YH81 brucella me
14	69	6.7	129	16 P71978	P71978 mycobacteri
15	68.5	6.7	148	16 Q97P37	Q97P37 streptococc
16	68.5	6.7	163	4 Q9H8Q1	Q9H8Q1 homo sapien

17	68	6.6	133	16 Q98J17	Q98J17 rhizobium 1
18	68	6.6	171	17 Q9UY30	Q9UY30 pyrococcus
19	67.5	6.6	151	16 O8YKF6	O8YKF6 anabena sp
20	67.5	6.6	193	16 O9RXR1	O9RXR1 deinococcus
21	67	6.5	153	16 Q92NX0	Q92NX0 rhizobium m
22	67	6.5	181	16 Q9UR36	Q9UR36 neisseria m
23	67	6.5	192	2 Q9ZGB7	Q9ZGB7 streptomyce
24	67	6.5	194	16 Q9RKK4	Q9RKK4 streptomyce
25	67	6.5	197	17 Q26850	Q26850 methanobact
26	66.5	6.5	151	16 Q9K3S9	Q9K3S9 streptomyce
27	66.5	6.5	166	16 Q9HWF2	Q9HWF2 pseudomonas
28	66.5	6.5	187	16 Q9CHE2	Q9CHE2 lactococcus
29	66.5	6.5	197	6 Q97673	Q97673 sus scrofa
30	66	6.4	138	16 Q9PD86	Q9PD86 xylella fas
31	66	6.4	171	4 Q96SA7	Q96SA7 homo sapien
32	66	6.4	140	16 Q91766	Q91766 pseudomonas
33	65.5	6.4	147	17 O58707	O58707 pyrococcus
34	65	6.3	153	16 Q9L099	Q9L099 streptomyce
35	65	6.3	170	6 Q28421	Q28421 gorilla gor
36	65	6.3	178	17 Q978C3	Q978C3 thermoplasm
37	65	6.3	188	2 Q9Z589	Q9Z589 zymomonas m
38	65	6.3	196	16 Q8UB23	Q8UB23 agrobacteri
39	64.5	6.3	189	16 Q98FR8	Q98FR8 rhizobium 1
40	64.5	6.3	190	17 Q9HWA3	Q9HWA3 halobacteri
41	64	6.2	129	2 Q44423	Q44423 agrobacteri
42	64	6.2	129	16 Q8U753	Q8U753 agrobacteri
43	64	6.2	131	5 Q95Y12	Q95Y12 bombyx mori
44	64	6.2	135	17 Q8TXL9	Q8TXL9 methanopyru
45	64	6.2	157	12 Q93115	Q93115 vaccinia vi

ALIGNMENTS

RESULT 1	Q29124	PRELIMINARY;	PRT;	191 AA.
AC	Q29124;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	VLA-2 (Fragment).			
OS	Sue scrofa (pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.			
OX	NCBI_TaxID=96823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=ENDOTHELIAL CELLS;			
RX	MEDLINE=95036279; PubMed=7949129;			
RA	Bahou W.F., Potter C.L., Mirza H.;			
RT	"The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific			
RT	recognition sequence for endothelial cell attachment and spreading;			
RL	Blood 84:3734-3741(1994).			
DR	EMBL: Z12137; CAA78125.1; -			
DR	HSSP: P17301; IAOX.			
DR	InterPro: IPR002035; VWF_A.			
DR	Pfam: PF00092; VWA; 1.			
DR	PRINTS: PRO0453; VWFADOMAIN.			
DR	SMART: SM00327; VWA; 1.			
DR	PROSITE: PS50234; VWFA; 1.			
FT	NON_TER			
FT	NON_TER			
FT	SEQUENCE 191 AA; 20896 MW; 2E228B472EC699D8 CRC64;			
SO	Q29124			
QY	42 MSELVTFRLRLVGLFLDPEDIQVGLVQYGESPVHWSLGDFFRTKEVYVRAKLSRREG 101			
DB	1 WDAVKNFLEKRVQGLDGLPTQVGLIYANNPVENVLNTFKTKAEVWVATSTTQGG 60			


```

RESULT 5
O8TWC9 PRELIMINARY; PRT; 108 AA.
ID O8TWC9;
AC O8TWC9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE Nitrogen regulatory protein PII homolog.
MK1106.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6134 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Poluehin N.N.,
RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natarale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Mal'kh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
DR EMBL; AE010400; AAM02319.1; -.
KW Complete proteome.
SQ SEQUENCE 108 AA; 12117 MW; 0678F55B81C69214 CRC64;

Query Match
Best Local Similarity 23.8%; Score 71.5; DB 17; Length 108;
Matches 25; Conservative 16; Mismatches 35; Indels 29; Gaps 5;

OY 55 KLFIDPEQ-----QVGL-----VOGESPVHWSLGDPR-TKEEVRAKXLSRE 100
Db 5 RLFVPDENGRVMMADVGVTGFAITFRGVAAPDMAGFELREDPESAIKALNDLSEK- 63
101 GRETTQAQIMVA-----CTEGFSQSHGGRPEARLLVVVTGDGE 139
OY 64 -----AVMIVTVPEECVEKLDKDAARLAGEYTTIVDVE 100

RESULT 6
O9RS46 PRELIMINARY; PRT; 164 AA.
ID O9RS46;
AC O9RS46;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Hypothetical protein DR2281.
GN DR2281.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=R1.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eison J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Croebly M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uitterlinden T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Frazer C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL; AE002060; AAF11833.1; -.
DR TIGR; DR2281; -.
DR InterPro; IPR001159; DS_RBD.
DR Pfam; PF00035; dsrm; 1.

```

```

DR SMART; SM00358; DSRM; 1.
DR PROSITE; PS50137; DS_RBD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 164 AA; 17528 MW; 3055ABA9D5A587F CRC64;

Query Match
Best Local Similarity 25.2%; Score 71; DB 16; Length 164;
Matches 38; Conservative 19; Mismatches 66; Indels 28; Gaps 7;

OY 36 SNSIYPMEVQTFRLVGLKFLFIDPEQIQVGLVGY-----GESPVH-----WSLG----- 81
Db 20 SRTLYPMNAKCDLRLV-----SLGLTPFEARHGPAPHRTRTHVKWSSGVYA 71
OY 82 --DERTKEEVRAAKNISRRGRETQTQAQIMVACTEGFSQSHGGRPEARLLVVVTGDGE 139
Db 72 TAEGRTKDAERLALAEALARE-LDGSDA.PVPAPAPATPIAQOSEWPPIYAQVLAEAVAA 130
OY 140 SHDGEELPALKAC--EAGRVTRYGIAYLVGH 168
Db 131 MEFARE-DATLDEVARDAGRFYRELLADLGH 160

RESULT 7
O58027 PRELIMINARY; PRT; 169 AA.
ID O58027;
AC O58027;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DE Hypothetical protein PH0289.
GN PH0289.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyma A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000001; BAA29361.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 169 AA; 19349 MW; 3C5FB65945CEFF7A CRC64;

Query Match
Best Local Similarity 6.9%; Score 71; DB 17; Length 169;
Matches 35; Conservative 14; Mismatches 45; Indels 42; Gaps 6;

OY 40 YPMSEVQTFRLVGLKFLFIDPEQIQVGLVGYGESP-----VHESLGD----- 83
Db 52 YSLSHVSSAMRVLG-----VGLVQVKKRGRKAVFIATKQFSEWRSSAFYEKI 101
OY 84 -----RTKEEVRAAKNISRRGRETQTQAQIMVACTEGFSQSHGGRPEARLLVVVTGD 137
Db 102 LRDIDEFRENLRLAKELHEHKGSEVERIKKKLLALK-----RNEVARKLLTLIMQ 153
OY 138 GESHDEELPALKAC 153
Db 154 FRSE--BELHLVLEKC 167

RESULT 8
O28796 PRELIMINARY; PRT; 170 AA.
ID O28796;
AC O28796;

```

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Complement C4 (Fragment)
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.
 NC NCB1_Taxid=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA O'Huigin C.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 RX MEDLINE=92104634; PubMed=1729169;
 RA Kawaguchi H., Zaleska-Ruczynska Z., Figueroa F., O'Huigin C.,
 RA Klein U.;
 RT "C4 genes of the chimpanzee, gorilla, and orang-utan: evidence for
 RT extensive homogenization."
 RL Immunogenetics 35:16-23(1992).
 DR EMBL; 231605; CAB3479.1; -.
 DR HSSP; P01024; IC3D.
 DR InterPro; IPR001559; Macrogloblina2.
 DR Pfam; PF00207; AZM; 1.
 FT NON_TER 1 1
 FT NON_TER 170 170
 SQ SEQUENCE 170 AA; 18045 MW; AE63A15A104B4854 CRC64;
 Query Match 6.9%; Score 71; DB 6; Length 170;
 Best Local Similarity 23.4%; Pred. No. 76;
 Matches 33; Conservative 19; Mismatches 79; Indels 10; Gaps 4;
 QY 34 DGSNSITPYMSEVQFLRLVGLFIDPEQIQVGLVQYGSFVHEWEL-----GDPRTKE 87
 DB 8 DGSYAMLSRDSSTWLTATLVKLV-LSLAQEQVGVSEPKQETSNTLSSQOQAGSDQDLS 66
 QY 88 EYVRAAKNLSRREGRETKTAQA-IMVACTEGFS--QSHGGRPEARLLVVTDSHDE 144
 DB 67 PVTHRGMGGLVGNDETVALTAFTVIALHHLGLAVFDQEGAEPLKQVEASISKANSFLGE 126
 QY 145 ELPAALKACEAGVTRYGIAV 165
 DB 127 KASAGLLGAHAAATYVALTL 147
 RESULT 9
 ID Q9X957 PRELIMINARY; PRT; 174 AA.
 AC Q9X957;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 19.0 kDa protein (Putative transcripional
 DE regulator).
 GN ORF4 OR SC02209 OR SC10B7.04.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 NC NCB1_Taxid=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=A;
 RA Vlierberg S., Frese D., Engels A., Mohlleben W.;
 RT "Analysis of the glnI region from Streptomyces coelicolor A3(2).";
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);

RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=97000351; PubMed=8843436;
 RA Redanbach M., Kleser H.M., Denaplatte D., Eichner A., Cullum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; Y13833; CAB43948.1; -.
 DR EMBL; AL355752; CAB90857.1; -.
 DR InterPro; IPR000835; HTH_MARR.
 DR PRINTS; PR00598; HTHMARR.
 DR SMART; SM00347; HTH_MARR; 1.
 KW DNA-binding; Hypothetical protein; Transcription regulation.
 SQ SEQUENCE 174 AA; 18985 MW; 58A15731E5C035 CRC64;
 Query Match 6.9%; Score 71; DB 16; Length 174;
 Best Local Similarity 31.4%; Pred. No. 78;
 Matches 27; Conservative 11; Mismatches 28; Indels 20; Gaps 4;
 QY 118 FQSHGGRPEARLLVVTD-----GESHDGELPAALKACAG-----RTRRGIAVLG 167
 DB 44 FAORNGMHPDVRLALIMDARAGEAATTAGHGLGALGNSAGTVALVRLERAG----- 98
 QY 168 HYLRRQRPDSFLEIRITIASDPPER 193
 DB 99 HVRVRDE----RDRRVTVVDER 119
 RESULT 10
 ID Q9M0V0 PRELIMINARY; PRT; 197 AA.
 AC Q9M0V0;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 21.8 kDa protein (MFDX1).
 GN ATG05450 OR ATMPDX1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NC NCB1_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vli D.M.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dechta N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Takubo K., Nonaka Y., Mizutani S., Takenaka S., Takahashi M.,
 RA Tsuyama S., Ohta D.,
 RT "Identification and characterization of a mitochondrial electron
 RT transfer chain in plants comprising of an adrenocortical ferredoxin
 RT homologue and its oxidoreductase."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
 RA Shin P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full length cDNA of gene At4g05450 (GI:15235586)."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shin P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.,
 RT "Arabidopsis Open Reading Frame (ORF) Clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN
 RP
 RA EMBL: AL161503; CAB81087.1; -
 RA EMBL: AB075738; BAB78226.1; -
 RA EMBL: AY065280; AAL38756.1; -
 RA HSSP: P00257; IAYF.
 DR InterPro: IPR001055; Adrenodoxin.
 DR InterPro: IPR001041; Ferredoxin.
 DR Pfam: PF00111; fer2; 1.
 DR PRINTS: PR00355; ADRENODOXIN.
 DR PROSITE: PS00814; ADX; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KM Hypothetical protein.
 SQ SEQUENCE 197 AA; 21830 MW; A35325C9F8C40A85 CRC64;
 Query Match 6.9%; Score 71; DB 10; Length 197;
 Best Local Similarity 22.7%; Pred. No. 93;
 Matches 27; Conservative 15; Mismatches 39; Indels 38; Gaps 5;
 QY 94 KNLRSREG-----RETKTAQAIWA-----CTEGFSQSHGCRPEARLLVVVTDG 138
 DB 29 KNLHSYGYLQSLPVPFRQARTSQEAMFLKSKCTSTTSSENGDEDETKITLIFVD- 87
 QY 139 ESHDEELPAALK-----ACEGRV--FTYGVAVGHYLRQRDS 177
 DB 88 -KDEELPVPKPIGMSVLEAAHENDIDLEGACSLACSTCHVIMPTETYNKLEP 144
 RESULT 11
 Q8ZME3 PRELIMINARY; PRT; 118 AA.
 AC Q8ZME3;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative transcriptional regulators containing the CopG/Arc/MetJ
 DE DNA-binding domain and a metal-binding domain.
 GN STM2955;
 OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L72 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.,
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2."
 RL Nature 413:852-856(2001).
 DR EMBL: AE008835; AAL21835.1; -
 DR InterPro: IPR005360; UPP0156.
 DR Pfam: PF03693; UPP0156; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 118 AA; 13575 MW; 35280AFAE671DAD CRC64;
 Query Match 6.9%; Score 70.5; DB 16; Length 118;
 Best Local Similarity 21.4%; Pred. No. 52;
 Matches 24; Conservative 15; Mismatches 24; Indels 49; Gaps 3;
 QY 81 GDFRTKEEVAAPKLSRREGRETKTAQAIWVACTEGFSQSHGCRPEARLLVVVTDGES 140
 DB 47 GDVYTSQVIRESLRLREKQAESR-LQALRELLAEGL--NSEBPQ----- 90
 QY 141 HDGEELPAALKACEAGVTRYGVAVGHYLRQRDPSSFLREITIASDPDE 192
 DB 91 -----WEKDAFLRKVTGMKIPDE 109
 RESULT 12
 Q8Z448 PRELIMINARY; PRT; 118 AA.
 AC Q8Z448;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein STY3093.
 GN STY3093.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahin A.M.,
 RA Baker S., Baughman D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jaseles K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.,
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18."
 RL Nature 413:848-852(2001).
 DR EMBL: AL627276; CAD06069.1; -
 DR InterPro: IPR005360; UPP0156.
 DR Pfam: PF03693; UPP0156; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 118 AA; 13606 MW; 43F36748B2164D59 CRC64;
 Query Match 6.9%; Score 70.5; DB 16; Length 118;
 Best Local Similarity 21.4%; Pred. No. 52;
 Matches 24; Conservative 15; Mismatches 24; Indels 49; Gaps 3;
 QY 81 GDFRTKEEVAAPKLSRREGRETKTAQAIWVACTEGFSQSHGCRPEARLLVVVTDGES 140

Db 47 GDRTOSEVIRESLRLLREKQASR-LOALRELLAGL---NSGEPOA----- 90
 QY 141 HOGELPALAKACEAGRVTRYGIAVLHRRORDSSFLREIRTIASDPDE 192
 Db 91 -----WEXDAFLRKXKTMIRKIDE 109

RESULT 13

OC8YHBI PRELIMINARY; PRT; 188 AA.

AC 08YHBI, 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
 DE Transcriptional regulator, TETR family.
 GN BME10891.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxId=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA DelVecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujar C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Resnik G.,
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elizer P.H., Hagius S., O'Callaghan D., Peterson J.-J.,
 RA Haselkorn R., Kyrides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AE009529; AAL52072.1; -
 DR InterPro: IPR001647; HTH_TecR.
 DR Pfam: PF00440; TecR, 1.
 DR PRINTS: PR00455; HTHTECR.
 KW Complete proteome.
 SQ SEQUENCE 188 AA; 20679 MW; B6F0102AD45104AA CRC64;

Query Match 6.8%; Score 70; DB 16; Length 188;
 Best Local Similarity 25.0%; Pred. No. 1.1e+02;
 Matches 28; Conservative 25; Mismatches 49; Indels 10; Gaps 5;

QY 85 TKEEVVRAKNSRREGRETKTAQAIMVACTEGFSQSGR-----PEARLLVVVDGSS 140
 Db 9 SREKIRATTELAQEVGPAHISDA--VAARAGLSK--GGLYSFPTKALLEANWEKYM 64
 QY 141 HOGELPALAKACEAGRVTRYGIAVLGHY-LRRORDSSFLREIRTIASDPD 191
 Db 65 GHEHQAQMAVOETLQSGERNRVARAFDVYRIQADKEPPA-CGVLLALAEHPD 115

RESULT 14

ID P71978 PRELIMINARY; PRT; 129 AA.

AC P71978; 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE Hypothetical protein RV1720C.
 GN RV1720C OR MTCY04C12.05C OR MT1761.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1173;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sutcliffe J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / OSHKOSH;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z81360; CAB03704.1; -
 DR EMBL: AE007037; AAK46033.1; -
 DR TIGR: MT1761; -
 DR TubercuList; RV1720C; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 129 AA; 13850 MW; D119173F733B06CD CRC64;

Query Match 6.7%; Score 69; DB 16; Length 129;
 Best Local Similarity 23.0%; Pred. No. 81;
 Matches 29; Conservative 19; Mismatches 50; Indels 28; Gaps 5;

QY 12 PQSLAPTAQR-----CPTVMDVIVLDGNSIYPSWEVQTLRLVGLFTDPDQIQ 64
 Db 17 PAG-AAVARLRRLRGETHVAPAHFDV-----EVIGAIRQAVVQLISDHEGL 60
 QY 65 VGLVOYGSPPVHNSLGDFTKEEVVRAKNSRREGRETKTAQAI--MACTEGFSQS 121
 Db 61 VVVVNFSLPVRWPLKPTORAYQLRSTHTVA--DGAVALAGLGVLITCGRLAQS 118
 QY 122 HGGRPE 127
 Db 119 HGHNRE 124

RESULT 15

ID Q97P37 PRELIMINARY; PRT; 148 AA.

AC Q97P37; 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
 DE Acetyltransferase, GNAI family.
 GN SP1807.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxId=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tetelini H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Peterson R.J.,
 RA Durkin A.S., Gwin M., Kolony J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser J.C.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 DR EMBL: AE007473; AAK75880.1; -
 DR TIGR: SP1807; -

DR Interpro; IPR001064; CRYSTALLIN.
DR Interpro; IPR001082; GCN5acetyltransf.
DR Pfam; PF00583; Acetyltransf. 1.
DR PROSITE; PS00225; CRYSTALLIN BETACAMMA; UNKNOWN_1.
KW Transferrase; Complete proteome.
SQ SEQUENCE 148 AA; 17286 MM; CCL0CA9A4A52634 CRC64;

Query Match	6.7%;	Score 68.5;	DB 16;	Length 148;
Best Local Similarity	25.0%;	Pred. No. 1.1e+02;		
Matches 28; Conservative	25;	Mismatches 34;	Indels 25;	Gaps 6;

```

Qy 9 SFGQGSGLAPPAQGCPTPYMDVIVLSDNSIYPMSEVQTRRL-----VGKFLDP 60
Db 27 SFGQGW-MGF---RIP-FLTLTQALBEVFSIPDQGEVPGVIGKIRLDELSDNLHGRFINP 81
Qy 61 EQLVGLVQVGESPVHMSLDDPFKE-----EYVAAKQLSRERE 103
Db 82 QKQKQGL---GSKALKRFVSLAEFNKRIDSLISLVFANPAAQNLVQKSEFE 130

```

Search completed: July 16, 2003, 08:45:01
Job time : 82 secs

THIS PAGE BLANK (USPTO)

PT Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain
XX
XX Example 2; Fig 5; 55pp; English.
XX
CC The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD1b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC the human integrin alpha subunit Alpha 10 A domain.
XX
SQ Sequence 195 AA;
Query Match 89.0%; Score 912; DB 23; Length 195;
Best Local Similarity 100.0%; Pred. No. 6.5e-97;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOVGLVOGESPVHWSLGD 82
DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOVGLVOGESPVHWSLGD 60
QY 83 FRTKEVVRRAKNLSRRREGRETQAQIMVACTEGFSQSHGGRPEARLLVVTDGESHD 142
DB 61 FRTKEVVRRAKNLSRRREGRETQAQIMVACTEGFSQSHGGRPEARLLVVTDGESHD 120
QY 143 GSELPAALKACBAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198
DB 121 GSELPAALKACBAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 176
RESULT 2
AAU76862 standard; Protein; 195 AA.
XX
AC AAU76862;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human integrin alpha subunit Alpha 10 variant A domain.
XX
KW Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
KW antiinflammatory; vasotropic; antiparasitic; vulnertary; gene therapy;
muten.
XX
OS Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 193
XX FT /note= "Wild-type Ile substituted by any other amino
XX FT acid"
XX
XX WO200209737-A1.
XX
XX PD 07-FEB-2002.
XX
XX PF 31-JUL-2001; 2001WO-US23957.
XX
XX PR 31-JUL-2000; 2000US-221950P.
XX
XX PR 11-JAN-2001; 2001US-0758493.

PR 13-MAR-2001; 2001US-0805354.
XX
XX (GENO) GEN HOSPITAL CORP.
XX
PI Arnaout AM, Li R, Xiong J;
XX
XX WPI; 2002-188687/24.
XX
PT Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain
XX
XX Claim 53; Page -; 55pp; English.
XX
CC The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD1b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC a human integrin alpha subunit Alpha 10 variant A domain.
CC Note: This variant sequence is not featured in the specification but has
CC been derived from the wild-type protein shown in AAU76853.
XX
SQ Sequence 195 AA;
Query Match 89.0%; Score 912; DB 23; Length 195;
Best Local Similarity 100.0%; Pred. No. 6.5e-97;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOVGLVOGESPVHWSLGD 82
DB 1 CPTVMDVIVLDGNSIYPMSEVQTFRLRVGKLFIDPEQIOVGLVOGESPVHWSLGD 60
QY 83 FRTKEVVRRAKNLSRRREGRETQAQIMVACTEGFSQSHGGRPEARLLVVTDGESHD 142
DB 61 FRTKEVVRRAKNLSRRREGRETQAQIMVACTEGFSQSHGGRPEARLLVVTDGESHD 120
QY 143 GSELPAALKACBAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198
DB 121 GSELPAALKACBAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 176
RESULT 3
AAU76854 standard; Protein; 193 AA.
XX
AC AAU76854;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human integrin alpha subunit Alpha 11 A domain.
XX
KW Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease;
KW antiinflammatory; vasotropic; antiparasitic; vulnertary; gene therapy.
XX
XX Homo sapiens.
XX
XX OS
XX PD 07-FEB-2002.
XX
XX PF 31-JUL-2001; 2001WO-US23957.

XX 31-JUL-2000; 2000US-221950P.
PR 11-JAN-2001; 2001US-0758493.
PR 13-MAR-2001; 2001US-0805354.
XX (GENO) GEN HOSPITAL CORP.
XX Arnaout AM, Li R, Xiong J;
XX WPI; 2002-188687/24.
XX
XX Novel high affinity integrin polypeptide useful for treating restenosis
XX and parasitic diseases, comprises all or part of variant integrin alpha
XX subunit A domain or variant integrin beta subunit A-like domain
XX
XX Example 2; Fig 5; 55pp; English.
XX
XX The invention relates to a high affinity integrin polypeptide comprising
XX all or part of a variant integrin alpha subunit A domain or a variant
XX integrin beta subunit A-like domain. The polypeptide, preferably the
XX CD11b alpha subunit A domain, where I at residue 332 has been replaced by
XX G or A, F at residue 313 and A at residue 320 have been replaced by C, or
XX V at residue 315 and A at residue 320 have been replaced by C, is useful
XX for determining if a test compound is a candidate compound for binding to
XX CD11b or for treating an inflammatory disorder, by contacting a test
XX compound with the polypeptide and determining if the test compound binds
XX to the polypeptide. The integrin subunits are useful for reducing
XX skeletal muscle injury, for treating disorders caused by
XX ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
XX diseases, to purify variant integrin polypeptide ligands and as bait
XX proteins in two-hybrid or three-hybrid assays. This sequence represents
XX the human integrin alpha subunit Alpha 11 A domain.
XX
SQ Sequence 193 AA;
Query Match 48.0%; Score 492; DB 23; Length 193;
Best Local Similarity 54.5%; Pred. No. 2.4e-48;
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;
QY 23 CPTWMDVIVLDGNSIYPMGEVOTFLRLVGKLFIDPEQIOVGIVOGESPVHEMSICD 82
Db 1 COTYMDIYIVLDGNSIYPMGEVOTFLRLVGKLFIDPEQIOVGIVOGESPVHEMSICD 60
QY 83 FRTKEEVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGGRPPAARLLVVTDEGSHD 142
Db 61 YRSVDVVEAASHIQRGTEFRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHD 118
QY 143 GEELPAALKACEAGVPTRYGIAVIGHYLRQRDPSSFLREIRTIASDDDERFFNV 198
Db 119 SPDLKVIQOSEBDVTRYAVAVLGYNRRGINPETFLNEIKYIASDPDKHFENV 174
RESULT 4
AAU76863
ID AAU76863 standard; Protein; 193 AA.
AC AAU76863;
XX
XX 21-MAY-2002 (first entry)
XX
XX Human integrin alpha subunit Alpha 11 variant A domain.
XX
XX Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;
XX A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
XX ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
XX antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
XX mutein.
XX Homo sapiens.
XX Synthetic.
XX Key Location/Qualifiers
XX Misc-difference 191

FT /note= "wild-type Ile substituted by any other amino
FT acid"
PN MO200209737-A1.
PD 07-FEB-2002.
XX 31-JUL-2001; 2001WO-US223957.
XX
XX 31-JUL-2000; 2000US-221950P.
PR 11-JAN-2001; 2001US-0758493.
PR 13-MAR-2001; 2001US-0805354.
XX (GENO) GEN HOSPITAL CORP.
XX Arnaout AM, Li R, Xiong J;
XX WPI; 2002-188687/24.
XX
XX Novel high affinity integrin polypeptide useful for treating restenosis
XX and parasitic diseases, comprises all or part of variant integrin alpha
XX subunit A domain or variant integrin beta subunit A-like domain
XX
XX Claim 57; Page -: 55pp; English.
XX
XX The invention relates to a high affinity integrin polypeptide comprising
XX all or part of a variant integrin alpha subunit A domain or a variant
XX integrin beta subunit A-like domain. The polypeptide, preferably the
XX CD11b alpha subunit A domain, where I at residue 332 has been replaced by
XX G or A, F at residue 313 and A at residue 320 have been replaced by C, or
XX V at residue 315 and A at residue 320 have been replaced by C, is useful
XX for determining if a test compound is a candidate compound for binding to
XX CD11b or for treating an inflammatory disorder, by contacting a test
XX compound with the polypeptide and determining if the test compound binds
XX to the polypeptide. The integrin subunits are useful for reducing
XX skeletal muscle injury, for treating disorders caused by
XX ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
XX diseases, to purify variant integrin polypeptide ligands and as bait
XX proteins in two-hybrid or three-hybrid assays. This sequence represents
XX a human integrin alpha subunit Alpha 11 variant A domain.
XX Note: This variant sequence is not featured in the specification but has
XX been derived from the wild-type protein shown in AAU76854.
XX
SQ Sequence 193 AA;
Query Match 48.0%; Score 492; DB 23; Length 193;
Best Local Similarity 54.5%; Pred. No. 2.4e-48;
Matches 96; Conservative 31; Mismatches 47; Indels 2; Gaps 1;
QY 23 CPTWMDVIVLDGNSIYPMGEVOTFLRLVGKLFIDPEQIOVGIVOGESPVHEMSICD 82
Db 1 COTYMDIYIVLDGNSIYPMGEVOTFLRLVGKLFIDPEQIOVGIVOGESPVHEMSICD 60
QY 83 FRTKEEVRAAKNLSRREGRETQTAQIMVACTEGFSQSHGGRPPAARLLVVTDEGSHD 142
Db 61 YRSVDVVEAASHIQRGTEFRTAFGIEFARSEAFQK--GGRKAKKVMIVITDGESHD 118
QY 143 GEELPAALKACEAGVPTRYGIAVIGHYLRQRDPSSFLREIRTIASDDDERFFNV 198
Db 119 SPDLKVIQOSEBDVTRYAVAVLGYNRRGINPETFLNEIKYIASDPDKHFENV 174
RESULT 5
AAU76851
ID AAU76851 standard; Protein; 195 AA.
AC AAU76851;
XX
XX 21-MAY-2002 (first entry)
XX
XX Human integrin alpha subunit Alpha 1 (CD49a) A domain.
XX
XX Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;

KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; Alpha 1;
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.
 OS Homo sapiens.
 XX WO200209737-A1.
 PN 07-FEB-2002.
 PD 31-JUL-2001; 2001WO-US23957.
 PF 31-JUL-2000; 2000US-221950P.
 PR 11-JAN-2001; 2001US-0758493.
 PR 13-MAR-2001; 2001US-0805354.
 XX (GEO) GEN HOSPITAL CORP.
 XX Arnaout AM, Li R, Xiong J;
 PI WPI; 2002-188687/24.
 DR Novel high affinity integrin polypeptide useful for treating restenosis
 PT and parasitic diseases, comprises all or part of variant integrin alpha
 PT subunit A domain or variant integrin beta subunit A-like domain -
 XX Example 2; Fig 5; 55pp; English.
 PS The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD1b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC the human integrin alpha subunit Alpha 1 (CD49a) A domain.
 SQ Sequence 195 AA;
 Query Match 40.7%; Score 417; DB 23; Length 195;
 Best Local Similarity 45.5%; Pred. No. 1.1e-39;
 Matches 80; Conservative 34; Mismatches 62; Indels 0; Gaps 0;
 QY 23 CPTMDVIVLDGNSIYPMSEVQTLRLVGLKFLFIDPEQIQVGLVQGESPVHWSLGD 82
 DB 1 CSTQDIIIVLDGNSIYPMDSVTAFLNLDLKRMDIGPKQTOVGIVQGENVTHFNLNK 60
 QY 83 FRTKEVRAAKNLSRREGRETAKAIVACTEGFSGHGGRPEAALLVVTGDESHD 142
 DB 61 YSSTEVEVLAARKIVQGRQMTIALGTDTARKAFTFARAGRVKVMVAVITDGESEHD 120
 QY 143 GEELPAALACAGAGRTRYGIAVLGHYLRDRDPSSFLAEIRTIADPDERFFENV 198
 DB 121 NHRLLKVIQDCEDENLQRFSTAILIGSYNRGNLSTEFVEELKSIASEPTEKHFENV 176
 RESULT 6
 AAU76860 standard; Protein; 195 AA.
 XX AAU76860;
 AC 21-MAY-2002 (first entry)
 XX Human integrin alpha subunit Alpha 1 (CD49a) variant A domain.
 DE
 XX

KW Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
 KW mutein; Alpha 1.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200209737-A1.
 PN 07-FEB-2002.
 PD 31-JUL-2001; 2001WO-US23957.
 PF 31-JUL-2000; 2000US-221950P.
 PR 11-JAN-2001; 2001US-0758493.
 PR 13-MAR-2001; 2001US-0805354.
 XX (GEO) GEN HOSPITAL CORP.
 XX Arnaout AM, Li R, Xiong J;
 PI WPI; 2002-188687/24.
 DR Novel high affinity integrin polypeptide useful for treating restenosis
 PT and parasitic diseases, comprises all or part of variant integrin alpha
 PT subunit A domain or variant integrin beta subunit A-like domain -
 XX Claim 45; Page -; 55pp; English.
 PS The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD1b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC a human integrin alpha subunit Alpha 1 (CD49a) variant A domain.
 CC Note: This variant sequence is not featured in the specification but has
 CC been derived from the wild-type protein shown in AAU76851.
 SQ Sequence 195 AA;
 Query Match 40.7%; Score 417; DB 23; Length 195;
 Best Local Similarity 45.5%; Pred. No. 1.1e-39;
 Matches 80; Conservative 34; Mismatches 62; Indels 0; Gaps 0;
 QY 23 CPTMDVIVLDGNSIYPMSEVQTLRLVGLKFLFIDPEQIQVGLVQGESPVHWSLGD 82
 DB 1 CSTQDIIIVLDGNSIYPMDSVTAFLNLDLKRMDIGPKQTOVGIVQGENVTHFNLNK 60
 QY 83 FRTKEVRAAKNLSRREGRETAKAIVACTEGFSGHGGRPEAALLVVTGDESHD 142
 DB 61 YSSTEVEVLAARKIVQGRQMTIALGTDTARKAFTFARAGRVKVMVAVITDGESEHD 120
 QY 143 GEELPAALACAGAGRTRYGIAVLGHYLRDRDPSSFLAEIRTIADPDERFFENV 198
 DB 121 NHRLLKVIQDCEDENLQRFSTAILIGSYNRGNLSTEFVEELKSIASEPTEKHFENV 176

Db 1 CSRLIDVVVCDSESNISYPMDAVKNFLKFKVGLDIPKTKVYGLIQVANNRPVENLNT 60
QY 83 FRKFEVVAANKLRSREGRETKTAQAIWVACTEGFSQSGRPEARLLVVTDGESH 142
Db 61 YKKEEMVIAVTSOTSGYGDLTNIPGAIQYARKYVSAASGGRRSATKVMVVVTDGKSHD 120
QY 143 GEEPLPAALACAGRYTRGIAVLGHYLRORDPSSFLREIKTIASDPDERFFNV 198
121 GSWLKAVIQCNHNDILRFGIALVGLNNALDPTKLIKIKAIASIPTERYFNV 176
RESULT 9
AAU09125
ID AAU09125 standard; protein, 185 AA.
XX AAU09125;
XX 16-JAN-2002 (first entry)
XX Human integrin alpha2 I domain.
XX DE
XX Human; integrin alpha2; I domain; protein co-ordinate data; thrombolytic;
KW cardiatic; cerebroprotective; hepatocytic; antiinflammatory; virucide;
KW vasotrophic; antiarteriosclerotic; thrombotic disorder;
KW myocardial infarction; stroke; acute thrombosis; angioplasty;
KW coronary bypass grafting; liver fibrosis; liver necrosis; hepatitis;
KW arterial occlusion; restenosis; atherogenesis; anti-platelet therapy.
XX OS
XX Homo sapiens.
XX Key Location/Qualifiers
FH Modified-site 17
FT /label= OTHER
FT /note= "Other= Unknown, represented by CPR in the
FT specification"
XX WO200173444-A2.
XX PN
XX 04-OCT-2001.
XX PD
XX 27-MAR-2001; 2001WO-GB01358.
XX PF
XX 28-MAR-2000; 2000US-192180P.
XX PR
XX (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.
XX PA
XX Farndale RW, Emsley J, Knight CG, Barnes MJ, Liddington RC;
XX PI
XX WPI; 2001-648466/74.
XX DR
XX New methods for identifying inhibitors of an I-domain-containing
PT polypeptide, particularly integrin, comprises use of coordinates of
PT peptide/receptor crystal structure -
XX PS
XX Example; Page 47; 100pp; English.
XX The invention relates to methods for identifying a potential inhibitor of
CC an I-domain comprising polypeptide, particularly integrin, comprising
CC using a three dimensional structure of the integrin alpha2 I-domain as
CC defined in the specification to design or select a potential inhibitor.
CC identified inhibitors are used to treat a disorder or disease. Such
CC disorders include thrombotic disorders, myocardial infarction and stroke,
CC acute thrombosis associated with angioplasty and coronary bypass
CC grafting, and with liver fibrosis or thrombotic complication of liver
CC necrosis after hepatitis. Inhibition of platelet alpha2beta1 may be used
CC to treat longer term occlusion of arteries, restenosis and atherogenesis.
CC Collagen receptor antagonism may be used as anti-platelet therapy.
CC The present sequence is the integrin alpha2 I domain.
XX SO
XX Sequence 185 AA;
XX Query Match 36.5%; Score 374; DB 22; Length 185;
XX Best Local Similarity 45.3%; Pred. No. 9.8e-35;

Matches 78; Conservative 28; Mismatches 66; Indels 0; Gaps 0;
QY 27 MDVVVLVGLNSNISYPMSEVQFLRLVGLKFLFIDPEQIOVGLVOGESPVHEWSLGDPTK 86
Db 3 IDVVVVCBESNISYKMDAVKNFLKFKVGLDIPKTKVYGLIQVANNRPVFNLTYYTK 62
QY 87 EEVVAANKLRSREGRETKTAQAIWVACTEGFSQSGRPEARLLVVTDGESH 146
Db 63 EEMVIAVTSOTSGYGDLTNIPGAIQYARKYVSAASGGRRSATKVMVVVTDGESH 122
QY 147 PAALACAGRYTRGIAVLGHYLRORDPSSFLREIKTIASDPDERFFNV 198
Db 123 KAVIDQCNHNDILRFGIALVGLNNALDPTKLIKIKAIASIPTERYFNV 174
RESULT 10
AAU19634
ID AAU19634 standard; Protein, 148 AA.
XX AAU19634;
XX 04-DEC-2001 (first entry)
XX DE
XX Human novel extracellular matrix protein, Seq ID No 284.
XX Human; secreted extracellular matrix protein; immunomodulatory;
KW Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiatic; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cyostatic;
KW antialzheimer's; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorders;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisenese; food additive.
XX OS
XX Homo sapiens.
XX PN
XX WO200155368-A1.
XX PD
XX 02-AUG-2001.
XX PF
XX 17-JAN-2001; 2001WO-US01348.
XX PR
XX 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.

PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465572/50.
DR N-PsDB; AAS31365.
XX
XX Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
PT
XX
XX Claim 11; SEQ ID No 444; 577bp; English.
XX
XX The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation.

Query Match 27 9%; Score 285.5; DB 22; Length 148;
Best Local Similarity 51.3%; Pred. No. 1.2e-24;
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;

Qy 1 GICARVDASFPQGSFLAFTACRCPYMDVIVLDGNSIYSEWVQFFLRLLVGKLFIDP 60
Db 21 GICSDVSPTFPQVNSIAIP-VQECSTQLDIVIVLDGNSIYPMWDSVTAFLNLLRRMDICP 79
Qy 61 FOIQVGLVQGESFPHEMSLGDFTKEVVAAAKLSRERERKTAQAIWVA 113
Db 80 KOTVGIVQYGENVTHEFNLNKYSTBEVLVAARKIVQRGGRQMTALGIDTA 132

RESULT 12
ABP47854
ID ABP47854 standard; Protein, 148 AA.
XX
XX ABP47854;
XX
XX 23-AUG-2002 (first entry)
DT
XX
XX Human polypeptide SEQ ID NO 284.
DE
XX
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilicer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
OS Homo sapiens.
XX
XX US2002042386-A1.
XX
XX 11-APR-2002.
PD
XX
XX 17-JAN-2001; 2001US-0764870.
PF
XX
XX 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 11-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217486P.
PR 14-JUL-2000; 2000US-217496P.
PR 16-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226868P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234977P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.

PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251866P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI: 2002-470713/50.
DR N-PSDB; ABQ66529.
XX
PT New nucleic acid encoding human proteins, useful for diagnosis,
PT treatment and prevention of e.g. osteoporosis, also related
PT polypeptides and antibodies -
XX
PS Claim 11; SEQ ID NO 284; 235pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
CC (ABP47846-ABP4810) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from USPTO at segdata.uspto.gov/sequence.html?DocID=99909764870.
XX
SQ Sequence 148 AA;
Query Match 27.9%; Score 285.5; DB 23; Length 148;
Best Local Similarity 51.3%; Pred. No. 1,2e-24;
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;
QY 1 GTCARVDASFQSGSLAFTARQRCFTYMDVVIYDGSNSIYPMSEVOTFLRLVGLKFLDP 60
Db 21 GICSDVSPFQVYNSTIAP-VQECSTQLDIVIVDGSNSIYPMDSVTAFLINDLERMDIGP 79
QY 61 EGIQVGLVQGSPEVHMSLGDPRTEKEVVRAXKNSRREGPTKTAQAIMYA 113
Db 80 KOTQVGIQVQGVNTHFNLNKYSTEEVLVAAKKIVQGRGTMTALGIDTA 132
RESULT 13
ABP48014
ID ABP48014 standard; Protein; 148 AA.
XX
XX ABP48014;
AC
XX
XX
DT 23-AUG-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 444.
XX
XX Human; nocrotropic; neuroprotective; cyostatic; dermatological; virucide;

KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
OS Homo sapiens.
XX
XX US2002042386-A1.
PN
PD 11-APR-2002.
XX
XX 17-JAN-2001; 2001US-0764870.
PF
XX 31-JAN-2000; 2000US-179965P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214866P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226688P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235634P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
PR 08-DEC-2000; 2000US-251866P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
XX Rosen CA, Ruben SM, Barash SC;
PI
XX WPI: 2002-470713/50.

DR N-PSDB; AB066689.
XX New nucleic acid encoding human proteins, useful for diagnosis,
PT treatment and prevention of e.g. osteoporosis, also related
PT polypeptides and antibodies
XX
PS Claim 11; SEQ ID NO 444; 235pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AB066521-AB066795) and proteins
CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=99909764870.
XX
SQ Sequence 148 AA:

Query Match 27.9%; Score 285.5; DB 23; Length 148;
Best Local Similarity 51.3%; Pred. No. 1.2e-24;
Matches 58; Conservative 19; Mismatches 35; Indels 1; Gaps 1;

OY 1 GICARVDSFOPGSLAPFAPRCPTFMDVIVLDGNSNTPYSEVQTFRLVGLFIDP 60
DB 21 GICSVSPFQVNSIAP-VQECSTQDLIVLDGNSNTPYSDSTAFINLLEMDIGP 79
OY 61 EQLQVGLVQGESPVHEWSLGDFTKEEVRAKNLSRREGRETYQAIVVA 113
DB 80 KOTGVGIQVGENVTHFPLNKYSTEEVLVAKKIVRGQGWTATGIDTA 132

RESULT 14
AAU87675
ID AAU87675 standard; Protein; 103 AA.
XX
AC AAU87675;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #585.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; lechemia; angio genesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234222.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-581633/65.
 XX N-PSDB; ABR44005.
 PT
 PT New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -
 XX

PS Claim 9; SEQ ID No 1193; 837bp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein (I) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC
 Query Match 27.5%; Score 282; DB 22; Length 103;
 Best Local Similarity 57.1%; Pred. No. 1.8e-24;
 Matches 56; Conservative 14; Mismatches 26; Indels 0; Gaps 0;
 QY 16 LAPTAORCPYMDVIVLDGNSIYPMSEVOTFRLRLGKLFIDPEQIQVGLVOYGESPV 75
 DB 1 LCPHPGCGQTYMDIYIVLDGNSIYPMSEVOTFRLRLGKLFIDPEQIQVGLVOYGESPV 60
 QY 76 HEMSLGDPRTKEEYVRAAKNISREGRRTKTAQALMVA 113
 DB 61 HEFLINDYRSVAKDYVEASHIEGRGTETRTAFGIXFA 98
 RESULT 15
 ID AAU19822 standard; Protein; 103 AA.
 XX
 AC AAU19822;
 XX
 DT 06-DEC-2001 (first entry)
 XX
 DE Human novel extracellular matrix protein, Seq ID No 472.
 XX
 KW Human; secreted extracellular matrix protein; immunomodulatory;
 KW Anti-HIV; antianemic; antithrombotic; antisclerotic; cardiac; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
 KW antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
 KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
 KW cancers; hyperproliferative disorder; breast neoplasia; melanoma;
 KW Sezary syndrome; Gaucher's disease; neurological diseases;
 KW Alzheimer's disease; Parkinson's disease; cardiovascular disorders;
 KW cardiac arrest; tachycardia; angina; infection; corneal infections;
 KW wound healing; immunogen; gene therapy; antisense; food additive.
 XX
 OS Homo sapiens.
 XX
 PN MO20015368-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01348.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0220863.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-465572/50.
N-PSDB; AAS31393.

Nucleic acid molecules encoding human secreted extracellular matrix
proteins, used in preventing, treating or ameliorating a disorder, e.g.
Alzheimer's and Parkinson's diseases and cancers -
Claim 11; SEQ ID No 472; 577bp; English.
The invention relates to isolated nucleic acid molecules encoding

CC novel human secreted extracellular matrix proteins (SPs). The
 CC polynucleotides and proteins are used to prevent, treat a medical
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
 CC chickens or sheep. For example, disorders associated with decreased
 CC expression of SPs. The SP polynucleotide or a vector expressing them may
 CC be administered to treat diseases by gene therapy. Antisense molecules
 CC may be administered to down regulate expression of SPs by binding with
 CC the cells own genes and preventing their expression. The polynucleotides
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be
 CC used as antigens to produce antibodies and to identify modulators
 CC (agonists and antagonists) of the SPs. The anti- (SP) antibodies and
 CC antagonists may also be used to down regulate expression and activity of
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.
 CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses
 CC include wound healing, maintenance of organs before transplantation,

Query Match 27.5%; Score 282; DB 22; Length 103;

Best Local Similarity 57.1%; Pred. No. 1.8e-24;

Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

```
QY 16 LAPTAQRCPYMDVNVIVLDGNSIYPWSEVQFLRLVGLFIDPEQIOVGLVQGESPV 75
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 LCPHFGCQOTMDIYIVLDGNSIYPWSEVQFLRLVGLFIDPEQIOVGLVQGESDVV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 76 HEWSLDGFRTEKEVVRANKLSRREGRETKTAQAINVA 113
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 HEFHLNDYRSVKQVVEASHIEGRGTEFRTAFGIKFA 98
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 16, 2003, 08:43:03
 Job time : 72 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:45:06 ; Search time 53 Seconds
(without alignments)
443.669 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337
Perfect score: 1025
Sequence: 1 GICARVDSFQPGSLAPTA.....FLREIRTIASDPDERFFENV 198

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues
Total number of hits satisfying chosen parameters: 232099

Minimum DB seq length: 0
Maximum DB seq length: 198

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	89.0	195	US-09-805-354-7	Sequence 7, Appl1
2	912	89.0	195	US-10-144-259-7	Sequence 7, Appl1
3	492	48.0	193	US-09-805-354-8	Sequence 8, Appl1
4	492	48.0	193	US-10-144-259-8	Sequence 8, Appl1
5	427	41.7	195	US-09-805-354-5	Sequence 5, Appl1
6	427	41.7	195	US-10-144-259-5	Sequence 5, Appl1
7	399	38.9	195	US-09-805-354-6	Sequence 6, Appl1
8	399	38.9	195	US-10-144-259-6	Sequence 6, Appl1
9	285.5	27.9	148	US-10-125-540-284	Sequence 284, App
10	285.5	27.9	148	US-10-125-540-444	Sequence 444, App
11	285.5	27.9	148	US-09-764-870-284	Sequence 284, App
12	285.5	27.9	148	US-09-764-870-444	Sequence 444, App
13	282	27.5	103	US-10-125-540-472	Sequence 472, App
14	282	27.5	103	US-09-764-870-472	Sequence 472, App
15	177.5	17.3	192	US-09-805-354-9	Sequence 9, Appl1
16	177.5	17.3	192	US-10-144-259-9	Sequence 9, Appl1
17	156	15.2	191	US-09-805-354-3	Sequence 3, Appl1
18	156	15.2	191	US-10-144-259-3	Sequence 3, Appl1
19	148	14.4	9	US-09-805-354-4	Sequence 4, Appl1

20	148	14.4	184	9	US-10-144-259-4	Sequence 4, Appl1
21	147	14.3	191	9	US-09-805-354-2	Sequence 2, Appl1
22	147	14.3	191	9	US-10-144-259-2	Sequence 2, Appl1
23	144	14.0	191	9	US-09-805-354-1	Sequence 1, Appl1
24	144	14.0	191	9	US-10-144-259-1	Sequence 1, Appl1
25	142	13.9	178	9	US-09-801-736A-4	Sequence 4, Appl1
26	140.5	13.7	185	9	US-10-143-819-18	Sequence 18, Appl1
27	135	13.2	171	10	US-09-864-761-33310	Sequence 33310, A
28	135	13.2	172	10	US-09-864-761-47637	Sequence 47637, A
29	123.5	12.0	193	9	US-09-984-245-293	Sequence 293, App
30	123.5	12.0	193	9	US-09-966-262-293	Sequence 293, App
31	123.5	12.0	193	9	US-09-983-966-293	Sequence 293, App
32	123.5	12.0	193	9	US-10-143-090-293	Sequence 293, App
33	111.5	10.9	160	9	US-09-801-761-3	Sequence 3, Appl1
34	91.5	8.9	159	9	US-09-864-761-49020	Sequence 49020, A
35	78	7.6	79	10	US-09-864-761-34986	Sequence 34986, A
36	73	7.1	196	9	US-10-156-761-10750	Sequence 10750, A
37	69	6.7	174	10	US-09-975-901-18	Sequence 18, Appl1
38	69	6.7	181	10	US-09-975-901-5	Sequence 5, Appl1
39	69	6.7	191	9	US-10-156-761-8978	Sequence 8978, App
40	67.5	6.6	124	9	US-09-796-692-1620	Sequence 1620, App
41	67.5	6.6	124	9	US-09-796-692-2496	Sequence 2496, App
42	67.5	6.6	124	9	US-10-040-862-1620	Sequence 1620, App
43	67.5	6.6	124	9	US-10-040-862-2496	Sequence 2496, App
44	67	6.5	54	10	US-09-864-761-49014	Sequence 49014, A
45	67	6.5	133	10	US-09-867-550-598	Sequence 598, App

ALIGNMENTS

RESULT 1
US-09-805-354-7
; Sequence 7, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amin
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; PRIOR FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 7
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-7

Query Match 89.0%; Score 912; DB 9; Length 195;
Best Local Similarity 100.0%; Pred. No. 1e-89;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 CPTMDVIVVDGNSIYPMSEVOTFLRLIVGKLFIDBEQIOVGIVQGESPVHEWSIGD 82
DB 1 CPTMDVIVVDGNSIYPMSEVOTFLRLIVGKLFIDBEQIOVGIVQGESPVHEWSIGD 60
QY 83 FRTSEVVRRAKNLSRRREGRTKTAQALMVACTGFSQSHGGRPAALLVVMDDGSHD 142
DB 61 FRTSEVVRRAKNLSRRREGRTKTAQALMVACTGFSQSHGGRPAALLVVMDDGSHD 120
QY 143 GEELPAALKAQAEGRVTRVYGIAVLGHVLRORDDSSFLREIRTIASDPDERFFENV 198
DB 121 GEELPAALKAQAEGRVTRVYGIAVLGHVLRORDDSSFLREIRTIASDPDERFFENV 176

RESULT 2

```

US-10-144-259-7
? Sequence 7, Application US/10144259
? Publication No. US20030109691A1
? GENERAL INFORMATION:
? APPLICANT: Arnaout, M. Amin
? APPLICANT: Li, Rui
? APPLICANT: Xiong, Jian-Ping
? TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: 00786-548001
? CURRENT APPLICATION NUMBER: US/10/144,259
? CURRENT FILING DATE: 2002-09-04
? PRIOR APPLICATION NUMBER: US 09/758,493
? PRIOR FILING DATE: 2001-01-11
? PRIOR APPLICATION NUMBER: US 60/221,950
? PRIOR FILING DATE: 2000-07-31
? NUMBER OF SEQ ID NOS: 30
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 7
? LENGTH: 195
? TYPE: prt
? ORGANISM: Homo sapiens
US-10-144-259-7

```

	Query Match	89.0%	Score 912	DB 9	Length 195	
	Best Local Similarity	100.0%	Pred. No. 1e-69			
	Matches 176	Conservative 0	Mismatches 0	Indels 0	Gaps 0	
Qy	23	CPTVMDVIVLDGNSIYPWSEVQFLRLRYGLFLIDPEQIQVGLQVYGESPVHWSLSD	82			
Db	1	CPTVMDVIVLDGNSIYPWSEVQFLRLRYGLFLIDPEQIQVGLQVYGESPVHWSLSD	60			
Qy	83	FRYKEEVVRAAKNLSRRREGRETKTQALIVACTEGFSQSHGRRPEARLLVVVTDSGSHD	142			
Db	61	FRYKEEVVRAAKNLSRRREGRETKTQALIVACTEGFSQSHGRRPEARLLVVVTDSGSHD	120			
Qy	143	GEELPAAKACAGVTRYGIVLGHVLRQDPDSFLFEIRITIASDPDEREPFNV	198			
Db	121	GEELPAAKACAGVTRYGIVLGHVLRQDPDSFLFEIRITIASDPDEREPFNV	176			

```

? RESULT 3
? US-09-805-354-8
? Sequence 8, Application US/09805354
? Publication NO. US20030078375A1 .
? GENERAL INFORMATION:
? APPLICANT: Arnaout, M. Amin
? APPLICANT: Li, Rui
? APPLICANT: Xiong, Jian-Ping
? TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: 00786-556001
? CURRENT APPLICATION NUMBER: US/09/805,354
? CURRENT FILING DATE: 2002-06-04
? PRIOR APPLICATION NUMBER: US 09/158,493
? PRIOR FILING DATE: 2001-01-11
? PRIOR APPLICATION NUMBER: US 60/221,950
? PRIOR FILING DATE: 2000-07-31
? NUMBER OF SEQ ID NOS: 20
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 8
? LENGTH: 193
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-805-354-8

```

	Query Match	Similarity	48.0%	Score	497	DB	9	Length	193
	Best Local	Similarity	54.5%	Pred. No.	9	8e-45			
	Matches	Conservative	31	Mismatches	47	Indels	2	Gaps	1
Oy	23	CPTWDDVIVYLDGNSISIPMSVQCFRLRYLTKFLIDPDLQIOGLVGYGESPYHNSLCD	:	:	:	:	:	:	82
			:	:	:	:	:	:	
Dd	1	CGTHMDIYLDGNSISIPMSVQCFRLRYLTKFLIDPDLQIOGLVGYGESDYVHEHNLND	:	:	:	:	:	:	60
Oy	83	FRTKEEVRAAKNLSRREGRETKTAQAIIWACTEGESOSHGRPEAARLLVVTVDGESH	:	:	:	:	:	:	142

[illegible]

```

RESULT 4
US-10-144-259-8
Sequence 8, Application US/10144259
Publication NO. US2003010961A1
GENERAL INFORMATION:
APPLICANT: Arnaout, M. Amin
APPLICANT: Li, Rui
APPLICANT: Xiong, Jian-Ping
TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 00786-548001
CURRENT APPLICATION NUMBER: US/10/144,259
CURRENT FILING DATE: 2002-09-04
PRIOR APPLICATION NUMBER: US 09/758,493
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: US 60/221,950
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ. ID NOS: 30
SOFTWARE: FastSeq for Windows Version 4.0
SEQ. ID NO. 8
LENGTH: 193
TYPE: PRT
ORGANISM: Homo sapiens
US-10-144-259-8

```

Query Match	48.0%;	Score 492;	DB 9;	Length 193;
Best Local Similarity	54.5%;	Pred. No. 9.8e-45;		
Matches	96;	Conservative	31;	Mismatches 47;
			Indels	2;
			Gaps	1

[illegible]

```

? RESULT 5
? US-09-805-354-5
? Sequence 5, Application US/09805354
? Publication No. US20030078375A1
? GENERAL INFORMATION:
? APPLICANT: Arnaout, M. Amin
? APPLICANT: Li, Rui
? APPLICANT: Xiong, Jian-Ping
? TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: 00786-536001
? CURRENT APPLICATION NUMBER: US/09/805,354
? CURRENT FILING DATE: 2002-06-04
? PRIOR APPLICATION NUMBER: US 09/758,493
? PRIOR FILING DATE: 2001-01-11
? PRIOR APPLICATION NUMBER: US 60/221,950
? PRIOR FILING DATE: 2000-07-31
? NUMBER OF SEQ ID NOS: 20
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 5
? LENGTH: 195
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-805-354-5

```


Best Local Similarity 57.1%; Pred. No. 1.3e-22;
Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Qy 16 LAPTORCPTVMDVIVLDGNSIYPMSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 75
Db 1 LCPHFGCQITWDIYIVLDGNSIYPMSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 60

Qy 76 HEMSLGDFRTKEEVVRAAKNLSRREGRETKTAQAIWVA 113
Db 61 HEFHLNDYRSVQDVVEAASHIEQRGCTETRTAFGIXFA 98

RESULT 14
US-09-764-870-472
Sequence 472, Application US/09764870
Patent No. US20020042386A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: P1214
CURRENT APPLICATION NUMBER: US/09/764,870
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 472
LENGTH: 103
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (96)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-472

Query Match 27.5%; Score 282; DB 10; Length 103;
Best Local Similarity 57.1%; Pred. No. 1.3e-22;
Matches 56; Conservative 14; Mismatches 28; Indels 0; Gaps 0;

Qy 16 LAPTORCPTVMDVIVLDGNSIYPMSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 75
Db 1 LCPHFGCQITWDIYIVLDGNSIYPMSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPV 60

Qy 76 HEMSLGDFRTKEEVVRAAKNLSRREGRETKTAQAIWVA 113
Db 61 HEFHLNDYRSVQDVVEAASHIEQRGCTETRTAFGIXFA 98

RESULT 15
US-09-805-354-9
Sequence 9, Application US/09805354
Patent No. US20030078375A1
GENERAL INFORMATION:

APPLICANT: Arnaout, M. Amin
APPLICANT: Li, Rui
APPLICANT: Xiong, Jian-Ping
TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
FILE REFERENCE: 00786-536G01
CURRENT APPLICATION NUMBER: US/09/805,354
CURRENT FILING DATE: 2002-06-04
PRIOR APPLICATION NUMBER: US 09/758,493
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: US 60/221,950
PRIOR FILING DATE: 2000-07-31
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-805-354-9

Query Match 17.3%; Score 177.5; DB 9; Length 192;

Best Local Similarity 31.0%; Pred. No. 4.8e-11;
Matches 54; Conservative 31; Mismatches 80; Indels 9; Gaps 5;

Qy 28 DVIIVLDGNSIYPMSEVOTFLRLVGLKLFIDPEQIOVGLVOYGESPVHEWSIGDPR 85
Db 6 EIALIIDSGSIDPPDDFORAKDIFSNMKNRYEKCFCENFALVOYGVIOTEPDLRSD 65

Qy 86 KEEVVRRAKNSRREGRETKTAQAIWVACTEFSQSHGGRPEARLIVVTDGE-SHDGE 144
Db 66 VMASLARVQNT-QVGSVTKTASAMQVLDISFTSSGSRRAKSKVMVVLTDGIFEDPL 124

Qy 145 ELPAALRACGKRVTRGIVLGHYLRQRPSSFLREIRITIASDPPDPRFFNV 198
Db 125 NLITVINSPKMGVVERFAIGV-GEFPRSART---ARELNLIASDPPDPRTHAFKV 173

Search completed: July 16, 2003, 08:51:47
Job time : 54 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:41:56 / Search time 27 Seconds
(without alignments)
215.768 Million cell updates/sec

Title: US-09-647-544-2_COPY_140_337

Perfect score: 1025

Sequence: 1 GICARVDAFQPGSLAFTA.....FLREIRIASDPDERFFENV 198

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 218932

Minimum DB seq length: 0

Maximum DB seq length: 198

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents, AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/6D.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	14.4	181	5	PCT-US96-01314-61
2	148	14.4	184	3	US-08-974-899-7
3	147	14.3	187	2	US-08-177-109A-62
4	147	14.3	187	2	US-08-687-706-62
5	147	14.3	187	5	PCT-US96-01314-60
6	145	14.1	184	3	US-08-974-899-8
7	144	14.0	187	2	US-08-177-109A-61
8	144	14.0	187	2	US-08-687-706-61
9	93	9.1	176	4	US-08-463-682-1
10	69	6.7	176	4	US-09-704-139-2
11	67.5	6.6	45	1	US-08-462-128-38
12	67.5	6.6	45	1	US-08-463-180-38
13	63.5	6.2	85	1	US-08-341-219-22
14	63.5	6.2	85	1	US-08-912-314A-22
15	60.5	5.9	75	4	US-08-864-357F-3
16	58.5	5.7	194	4	US-09-125-619-28
17	58	5.7	29	5	PCT-US96-01314-68
18	58	5.7	128	3	US-08-782-480-6
19	58	5.7	128	4	US-08-954-211-6
20	58	5.7	128	4	US-09-005-167A-6
21	57.5	5.6	128	4	US-09-176-741B-6
22	57.5	5.6	83	4	US-09-357-251-4
23	57.5	5.6	132	1	US-08-470-179-27
24	57.5	5.6	156	4	US-09-071-035-16
25	57.5	5.6	176	4	US-09-071-035-14
26	57.5	5.6	177	4	US-09-370-838-208
27	57.5	5.6	185	4	US-09-153-586-2

28	56.5	5.5	111	4	US-09-325-932A-63	Sequence 63, Appl
29	56.5	5.5	155	1	US-08-339-152A-22	Sequence 22, Appl
30	56.5	5.5	155	2	US-08-007-999B-7	Sequence 7, Appl
31	56.5	5.5	155	2	US-08-689-276A-7	Sequence 7, Appl
32	56	5.5	165	4	US-09-134-001C-5331	Sequence 5331, Ap
33	56	5.5	189	4	US-09-125-619-20	Sequence 20, Appl
34	56	5.5	195	1	US-08-438-753B-32	Sequence 32, Appl
35	56	5.5	195	1	US-08-443-883A-32	Sequence 32, Appl
36	56	5.5	195	2	US-08-631-328-32	Sequence 32, Appl
37	56	5.5	195	2	US-08-455-524B-32	Sequence 32, Appl
38	56	5.5	195	2	US-08-455-524B-32	Sequence 32, Appl
39	56	5.5	195	4	US-09-045-467-32	Sequence 32, Appl
40	55.5	5.4	46	1	US-08-462-128-40	Sequence 40, Appl
41	55.5	5.4	46	1	US-08-463-180-40	Sequence 40, Appl
42	55.5	5.4	99	2	US-08-795-444-1	Sequence 2, Appl
43	55.5	5.4	178	1	US-07-929-580B-2	Sequence 11, Appl
44	55.5	5.4	195	2	US-08-468-576B-11	Sequence 11, Appl
45	55.5	5.4	195	2	US-08-468-579B-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
PCT-US96-01314-61
; Sequence 61, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaut
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/01314
; FILING DATE: 30-JAN-96
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/380,167
; FILING DATE: 30-JAN-95
; ATTORNEY/AGENT INFORMATION:
; NAME: John W. Freeman
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00786/267001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 181 amino acids
; TYPE: amino acid
; STRANDEDNESS: linear
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-01314-61

Query Match 14.4%; Score 148; DB 5; Length 181;
Best Local Similarity 25.4%; Pred. No. 1.5e-10;
Matches 44; Conservative 37; Mismatches 70; Indels 22; Gaps 6;

QY 27 MDVIVVLDGNSITP--MSEVQFLRLRYGLFLDPEQIQGLVQYGESPFHMSLQGF- 83
DB 6 VDLVFLFDGMSLQPFDFQKILDFMKQVKKL--SNTSYOPAAVQFSTSYKTFDFDSYV 63
```

QY 84 ---RTKEVVRRAKNLSRREGRETKTAQAIMVACTEGESGSHGRPREARLLVVVDGESH 141
Db 64 KMKDPALAKVYKHMILL-----TNTGAINVYATEVFRRELGRPATKVLIIITDGEAT 119
QY 142 DGEELPALAKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERF 194
Db 120 DSGNIDAAMD-----IIRYIIIGIGHGFOTKESQET-----LHKFASKPASEF 161

RESULT 2

US-08-974-899-7
Sequence 7, Application US/08974899
Patent No. 6037454
GENERAL INFORMATION:
APPLICANT: Presta, Leonard G.
TITLE OF INVENTION: Humanized Anti-CD11a Antibodies
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,899
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031971
FILING DATE: 11/27/96
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P1014R1
TELEPHONE: 650/225-1994
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-974-899-7

Query Match 14.4%; Score 148; DB 3; Length 184;
Best Local Similarity 25.4%; Pred. No. 1.5e-10;
Matches 44; Conservative 37; Mismatches 70; Indels 22; Gaps 6;

QY 27 MOVVIVLDGNSIYP--MSEVOTFLRLVGLKLFIDEQIQVGLVGYGESPVHEWSIGDP- 83
Db 4 VDLVLFPSDSMSIQDPFQKIIDFMKDVKKL--SNTSIQFAAVQPSISYKTEFDFSDIV 61
QY 84 --RTKEVVRRAKNLSRREGRETKTAQAIMVACTEGESGSHGRPREARLLVVVDGESH 141
Db 62 KMKDPALAKVYKHMILL-----TNTGAINVYATEVFRRELGRPATKVLIIITDGEAT 117
QY 142 DGEELPALAKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERF 194
Db 118 DSGNIDAAMD-----IIRYIIIGIGHGFOTKESQET-----LHKFASKPASEF 159

RESULT 3;
US-08-177-109A-62
Sequence 62, Application US/08177109A
Patent No. 5869615

GENERAL INFORMATION:
APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby
TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/177,109A
FILING DATE: 03-JAN-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: WU 107
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-177-109A-62

Query Match 14.3%; Score 147; DB 2; Length 187;
Best Local Similarity 26.7%; Pred. No. 2.1e-10;
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

QY 23 CPTY-MDVIVLDGNSIYP--MSEVOTFLRLVGLKLFIDEQIQVGLVGYGESPVHEWS 79
Db 1 CPPOQDIFVLDGSGSISRPFATMNFVAIVISQ--STQFSIMQPSNKFQTHFT 58
QY 80 LGDRRTKEVVRRAKNLSRREGRETKTAQAIMVACTEGESGSHGRPREARLLVVVDGE 139
Db 59 FEEFRRTSNPLSLASVHQLOG--FTYTRAIQNVVHRLFHASYGARRADTKILIVTDGK 117
QY 140 SH-DGEELPALAKACEAGRVTRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFENV 198
Db 118 KEGDSLVDKVIDYIPMADAAGIRVAIGVLAQNRR-----NSWKELNDIASKPSQEHIFKV 172

RESULT 4
US-08-687-706-62
Sequence 62, Application US/08687706
Patent No. 5928892
GENERAL INFORMATION:
APPLICANT: Dennis E. Hourcade and Teresa J. Oglesby
TITLE OF INVENTION: MODIFIED COMPLEMENT PROTEASES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,706
FILING DATE: 26-JUL-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/177,109
FILING DATE: 03-JAN-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: WU 107 DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 62:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-687-706-62

Query Match 14.3%; Score 147; DB 2; Length 187;
Best Local Similarity 26.7%; Pred. No. 2.1e-10;
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

Qy 23 CPTV-MDVIVLDGNSIYP--WSEVQFLRLVGLFIDPEQIQVGLVQGESPVHMS 79
Db 1 CPRODQIVFLIDSGSISRNFATMNFVRAVISQ-FQRP-STQFSLMQSNKFOQHT 58
Qy 80 LGDPRTKEVVRRAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVVDGE 139
Db 59 FEEFRRTSNPLSLASVHQLQG-FTYATATQNVVHRLFHASVYCARDAKILIVTDGK 117
Qy 140 SH-DGEELPAALKACAGRVTVYGIAYVGHYLRQRDPSSFLEIRTIASDPDERFFPNV 198
Db 118 KEGSLDYKDVIPMDADAGIIRYALGVGLAFQNR-----NSWKELNDIASRPSQEHIFKV 172

RESULT 5
PCT-US96-01314-60
Sequence 60, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaout
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
TITLE OF INVENTION: ANTAGONISTS
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-01314-60

Query Match 14.3%; Score 147; DB 5; Length 187;
Best Local Similarity 26.7%; Pred. No. 2.1e-10;
Matches 48; Conservative 36; Mismatches 84; Indels 12; Gaps 7;

Qy 23 CPTV-MDVIVLDGNSIYP--WSEVQFLRLVGLFIDPEQIQVGLVQGESPVHMS 79
Db 1 CPRODQIVFLIDSGSISRNFATMNFVRAVISQ-FQRP-STQFSLMQSNKFOQHT 58
Qy 80 LGDPRTKEVVRRAKNLSRREGRETKTAQAIMVACTEGFSQSHGGRPEARLLVVVDGE 139
Db 59 FEEFRRTSNPLSLASVHQLQG-FTYATATQNVVHRLFHASVYCARDAKILIVTDGK 117
Qy 140 SH-DGEELPAALKACAGRVTVYGIAYVGHYLRQRDPSSFLEIRTIASDPDERFFPNV 198
Db 118 KEGSLDYKDVIPMDADAGIIRYALGVGLAFQNR-----NSWKELNDIASRPSQEHIFKV 172

RESULT 6
US-08-974-899-8
Sequence 8, Application US/08974899
Patent No. 6037454
GENERAL INFORMATION:
APPLICANT: Prestea, Leonard G.
APPLICANT: Jardiou, Paula M.
TITLE OF INVENTION: Humanized Anti-CD11a Antibodies
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,899
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031971
FILING DATE: 11/27/96
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER: 40,378
REFERENCE/DOCKET NUMBER: P1014R1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1994
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 184 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-974-899-8

Query Match 14.1%; Score 145; DB 3; Length 184;


```
Sequence 1, Application US/08463682
Patent No. 6008193
GENERAL INFORMATION:
APPLICANT: Leonard Garfinkel, et al.
TITLE OF INVENTION: Cloning and Production of Human Von
Willebrand Factor Grib Binding Domain Polypeptides and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,682
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36537-B2-Y
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-463-682-1

Query Match
Best Local Similarity 9.1%; Score 93; DB 3; Length 187;
Matches 38; Conservative 46; Mismatches 72; Indels 28; Gaps 6;

Qy 23 CPTWMDVIVLDGSN--GIYPSWVQTFRLRLVGLFDPEIQVGLVQYGESPVHMSL 80
Db 1 CSRLIDLVFLLDGSSRLSEAEFEVKAFAVDMMERLRISQKWRVAVVEYHDGSHAYIGL 60
Qy 81 GDFRTKEVVRPAKULSRREGRETQIAQIMVACTEGFSQSHGSRPEARLLVVTDGDS 140
Db 61 KDRKPSRLRIASQVYKAGSVASTSEVLKTLFQITSKI--DPEPSRIALLL----- 113
Qy 141 HDGEELPALAKCEAGRVTRYGIIVLGHYLRQR-----DPSSFLREIRTTIASD-PD 191
Db 114 -----MASQEPQRMSRNVRYV-QGLKKKQIVIVPGIGPHANLKQIRLIKQKQBE 163
Qy 192 ERF 195
Db 164 NKA 167

RESULT 10
US-09-704-139-2
Sequence 2, Application US/09704139
Patent No. 6420153
GENERAL INFORMATION:
APPLICANT: Kapellier, Rosanna
TITLE OF INVENTION: 18232, A NOVEL DUAL SPECIFICITY PHOSPHATASE AND USES THEREFOR
FILE REFERENCE: 10448-018001
CURRENT APPLICATION NUMBER: US/09/704,139
CURRENT FILING DATE: 2000-11-01
```

```
PRIOR APPLICATION NUMBER: US 60/185,772
PRIOR FILING DATE: 2000-02-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 176
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)_(1390)
OTHER INFORMATION: n = a, t, g, or c
US-09-704-139-2

Query Match
Best Local Similarity 6.7%; Score 69; DB 4; Length 176;
Matches 37; Conservative 17; Mismatches 50; Indels 26; Gaps 7;

Qy 89 VRAAKNL---SRREGRTQAQIMVACTEGFSQSHGSRPEARLLVVTDGESH- 142
Db 19 LVRVAPSLFLGASABAQAEQIARAGVTLGVNSRQGPAPGVAELRVVFPDPAIDL 78
Qy 143 -----GEELPALKA-----CEAGRVTRYGIIVLGHYLRQRDPS-----SFLREI 183
Db 79 LAHLEPTCAAMEAAVRAAGACLVYCKNGR--SRSAAVCTAYLMHRIGLSLAKAFQWYKSA 136
Qy 184 RTIASDPDERPF 195
Db 137 RPVA-ERNPGFW 147

RESULT 11
US-08-462-128-38
Sequence 38, Application US/08462128
Patent No. 5686059
GENERAL INFORMATION:
APPLICANT: Goetlinck, Paul F.
APPLICANT: Tondravi, Mehرداد
TITLE OF INVENTION: CARTILAGE MATRIX PROTEIN AND METHODS FOR
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lathive & Cockfield
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,128
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION:
APPLICATION NUMBER: US 08/006,096
FILING DATE: 15-JAN-1993
APPLICATION NUMBER: US 07/866,403
FILING DATE: 10-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Myers, Paul L.
REGISTRATION NUMBER: 35,965
REFERENCE/DOCKET NUMBER: MGP-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 45 amino acids
TYPE: amino acid
```


COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/912,314A
FILING DATE: 30-JUN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/341,219
FILING DATE: 05-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 8399-003-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-912-314A-22

Query Match 6.2%; Score 63.5; DB 4; Length 85;
Best Local Similarity 32.9%; Pred. No. 2.2; Mismatches 26; Indels 19; Gaps 4;
Matches 25; Conservative 6;

Qy 110 IMVACTESFSQ--SHGGRPEARLLVVTGDSHDEBIPALAKCEAGRTYGIAYIG¹⁶⁷
Db 14 LCVGQLSFAQWHSWGYPGKREL-----DSFGTSEISEIKLCEAGECG-----59

Qy 168 HYLRQRDPSSFLREI 183
Db 60 -YLRPQR--RSLIRNI 72

RESULT 15
US-08-864-357F-3
Sequence 3, Application US/08864357F
Patent No. 6255281
GENERAL INFORMATION:
APPLICANT: Claragen, Inc. & NIH
TITLE OF INVENTION: Use of Recombinant Human Uteroglobin in Treatment of Inflammato
FILE REFERENCE: 116142/2
CURRENT APPLICATION NUMBER: US/08/864,357F
CURRENT FILING DATE: 1997-05-28
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin version 3.0
SEQ ID NO 3
LENGTH: 75
TYPE: PRT
ORGANISM: rat
US-08-864-357F-3

Query Match 5.9%; Score 60.5; DB 4; Length 75;
Best Local Similarity 32.4%; Pred. No. 4.4; Mismatches 17; Indels 19; Gaps 5;
Matches 24; Conservative 14;

Qy 23 CPTWMDVV-IVLDGNSIY-----PMSVQ--TFLRLVGLKLFIDPEQIOVGLVQY 70
Db 3 CPGLQVLEALLGSESNEYEALKFPNPASDLOANAGTQKRLVDTL--PQTRINIVKL 59

Qy 71 GE---SPVHEWSL 80

Db 60 TEKILTSPLCEQDL 73

Search completed: July 16, 2003, 08:46:20
Job time : 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 1.79677 Seconds
(without alignments)
1872.646 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHPHYGLPVGPGEFFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	65.5	34.1	301	2	nodp protein - Azo
2	54	28.1	2139	2	vitellinogenin - yel
3	52	27.1	101	2	hypothetical prote
4	52	27.1	437	1	transforming prote
5	52	27.1	564	2	gene gll protein -
6	52	27.1	649	2	C-eki protein FB27
7	52	27.1	713	2	C-eki protein FB28
8	52	27.1	728	1	transforming prote
9	52	27.1	750	2	C-eki protein FB29
10	51.5	26.8	390	2	hypothetical prote
11	51.5	26.8	507	1	cytochrome P450 52
12	51	26.6	402	2	hypothetical prote
13	51	26.6	890	2	NFAT transcription
14	51	26.6	1053	2	sucrose-phosphate
15	50.5	26.3	300	2	probable methylene
16	50.5	26.3	317	2	ATP sulfurylase, s
17	50	26.0	428	2	hypothetical prote
18	49.5	25.8	666	2	alcohol dehydrogen
19	49.5	25.8	668	2	glycosyl hydrolase
20	49	25.5	136	2	hypothetical prote
21	49	25.5	286	2	probable ribosomal
22	49	25.5	446	2	proline/betaine tr
23	49	25.5	753	2	dolichyl-1-phosphate
24	49	25.5	771	2	transposase - Yers
25	48.5	25.3	317	2	sulfate adenylyate
26	48.5	25.3	341	2	hypothetical prote
27	48.5	25.3	352	2	probable argc prot
28	48.5	25.3	505	2	probable tpr prote
29	48.5	25.3	653	2	probable tsbu prot

30	48.5	25.3	759	2	S53922	PM76 protein - yea
31	48.5	25.3	1046	2	B983324	probable rnd efflu
32	48.5	25.3	1046	2	AD29959	AcrtB/AcrD/AcrF fam
33	48.5	25.3	1185	2	T36761	probable DNA polym
34	48	25.0	263	2	T35686	phosphatidylglycer
35	48	25.0	368	2	S57273	lignin peroxidase
36	48	25.0	492	2	G70749	probable hycr prot
37	48	25.0	661	1	TNBE12	74K alpha trans-in
38	48	25.0	728	2	S59964	procollagen-lysine
39	48	25.0	899	2	B75018	hypothetical prote
40	48	25.0	1262	2	T30524	protein phosphatas
41	48	25.0	1289	2	T18212	parasporel crystal
42	47.5	24.7	129	2	D70531	hypothetical prote
43	47.5	24.7	255	2	T13437	acid phosphatase h
44	47.5	24.7	299	1	ZZZRN	3'-phosphadenosin
45	47.5	24.7	299	2	G95932	probable sulfate a

ALIGNMENTS

RESULT 1

139754
nodp protein - Azospirillum brasiliense

C:Species: Azospirillum brasiliense

C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 11-May-2000

C:Accession: 139754

R:Vielle, C.; Elmerich, C.

Mol. Plant Microbe Interact. 3, 389-400, 1990

A:Title: Characterization of two Azospirillum brasiliense Sp7 plasmid genes homologous to

A:Reference number: 139754; MUID:92033082; PMID:2131098

A:Accession: 139754

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-301 <RES>

A:Cross-references: GB:M54866; NID:G142423; PIDN:AAA22185.1; PID:G142424

A:Gene: nodp

C:Superfamily: nodulation protein nodp

Query Match

Best Local Similarity 34.1%; Score 65.5; DB 2; Length 301;

Matches 15; Conservative 5; Mismatches 8; Indels 3; Gaps 2;

Db 228 GRPLNPGETEMR-RVRFRTLCYPLSGAI 257

7 GTLPVPG--DEPKTLRVQNLGCVVSGLI 35

||||| : ||| : ||| : |||

RESULT 2

546404
vitellinogenin - yellow fever mosquito

C:Species: Aedes aegypti (yellow fever mosquito)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 21-Jul-2000

C:Accession: 546404

R:Chen, J.S.; Cho, W.L.; Raikhel, A.S.

J. Mol. Biol. 237, 641-647, 1994

A:Title: Analysis of mosquito vitellinogenin cDNA. Similarity with vertebrate phosphatins a

A:Reference number: 546404; MUID:94210487; PMID:8158643

A:Accession: 546404

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2139 <CHR>

A:Cross-references: EMBL:U02548; NID:G472307; PIDN:AAA18221.1; PID:G472308

Query Match

Best Local Similarity 28.1%; Score 54; DB 2; Length 2139;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 2 YEHPHYGLPVGPGEFFK 18

||||| : ||| : ||| : |||

366 DVHPDYKVPWGPADLK 382

```

A:Cross-references: GB:565038; NID:g410635; PID:g410636
C:Genetics:
A:Gene: gll1
C:Superfamily: gll transforming protein

Query Match          27.1%; Score 52; DB 2; Length 564;
Best Local Similarity 60.0%; Pred. No. 27;
Matches      9; Conservative    2; Mismatches     4; Indels      0; Gaps      0;

Oy       1 YEHPVGYTLPGVGPGP 15
           |||         |||
           :||:||||
Db        472 YEARGPSLPLGPGP 486

RESULT 6
C-ski protein FB27 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 18-Jun-1993
C:Accession: C32575
Mol. Cell. Biol. 9, 4046-4051, 1989
R:Sutcliffe, P.; Hughes, S.H.
A>Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski gene
A:Reference number: A32575; MUID:89384635; PMID:2779576
A:Accession: C32575
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-649 <SUT>
A:Cross-references: GB:M28517
C:Superfamily: ski transforming protein

Query Match          27.1%; Score 52; DB 2; Length 649;
Best Local Similarity 35.3%; Pred. No. 32;
Matches      12; Conservative    4; Mismatches     8; Indels     10; Gaps      1;

Oy       10 PVGGPPEF-----KTLRVQNLGCYVSG 33
           ||| | | | | | | | | | | | | | |
Db        67 PVMPGFPMPSDRSTERCITILEGTISCFTVGS 100

RESULT 7
C-ski protein FB28 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 18-Jun-1993
C:Accession: B32575
Mol. Cell. Biol. 9, 4046-4051, 1989
R:Sutcliffe, P.; Hughes, S.H.
A>Title: Isolation and characterization of three distinct cDNAs for the chicken c-ski gene
A:Reference number: A32575; MUID:89384635; PMID:2779576
A:Accession: B32575
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-713 <SUT>
A:Cross-references: GB:M28517
C:Superfamily: ski transforming protein

Query Match          27.1%; Score 52; DB 2; Length 713;
Best Local Similarity 35.3%; Pred. No. 35;
Matches      12; Conservative    4; Mismatches     8; Indels     10; Gaps      1;

Oy       10 PVGGPPEF-----KTLRVQNLGCYVSG 33
           ||| | | | | | | | | | | | | | |
Db        67 PVMPGFPMPSDRSTERCITILEGTISCFTVGS 100

RESULT 8
TVHDSK
transforming protein ski - human
C:Species: Homo sapiens (man)
C>Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 18-Jun-1999
C:Accession: S06053
C:Nomura, N.; Sasamoto, S.; Ishii, S.; Date, T.; Matsui, M.; Ishizaki, R.
```


NFAT transcription factor phosphoprotein - mouse (fragment)
 N:Alternate names: nuclear factor of activated T cells, phosphoprotein chain
 C:Species: Mus musculus (house mouse)
 C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
 C:Accession: A48753
 R:McCauley, P.G.; Luo, C.; Karpola, T.K.; Jain, J.; Badalian, T.M.; Ho, A.M.; Burgeon, Science 262, 750-754, 1993
 A>Title: Isolation of the cyclosporin-sensitive T cell transcription factor NFATp.
 A:Reference number: A48753; PMID:94053710; PMID:8235597
 A:Accession: A48753
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-890 <MCC>
 A:Cross-references: GB:U02079
 C:Keywords: phosphoprotein; transcription factor

Query Match 26.6%; Score 51; DB 2; Length 890;
 Best Local Similarity 75.0%; Pred. No. 62;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPYGTLPVGRGP 15
 DB 805 HPGLTCVLPGR 816

RESULT 14
 S34172
 sucrose-phosphate synthase (EC 2.4.1.14) - potato
 C:Species: Solanum tuberosum (potato)
 C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
 C:Accession: S34172
 R:Sonnwald, U.; Baener, A.
 submitted to the EMBL Data Library, June 1993
 A:Reference number: S34172
 A:Accession: S34172
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1053 <SON>
 A:Cross-references: EMBL:X73477; NID:g313264; PIDN:CAA51872.1; PID:g313265
 C:Function:
 A:Description: catalyzes the formation of sucrose-6-phosphate from UDPglucose and D-fruc
 A:Pathway: sucrose biosynthesis
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
 F:168-651/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

Query Match 26.6%; Score 51; DB 2; Length 1053;
 Best Local Similarity 35.5%; Pred. No. 75;
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPYGTLPVGRGPFTTLRVONLCVV 31
 DB 911 FKVCKRGTVF--PSKELRKVMRIQALRCNAV 939

RESULT 15
 T03464
 probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) / methylenetetrahy
 C:Species: Rhodobacter capsulatus
 C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 03-Jun-2002
 C:Accession: T03464
 R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
 Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
 A>Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003
 A:Reference number: Z14955; PMID:97404404; PMID:9256491
 A:Accession: T03464
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-300 <VIC>
 A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16117.1; PID:g3128265
 C:Genetics:
 A:Map position: 1
 C:Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

C:Keywords: hydrolase; multifunctional enzyme; oxidoreductase
 F:9-287/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

Query Match 26.3%; Score 50.5; DB 2; Length 300;
 Best Local Similarity 38.5%; Pred. No. 22;
 Matches 10; Conservative 5; Mismatches 8; Indels 3; Gaps 1;

QY 2 EVHPYGT--TLVGRGPFTTLRVQ 24
 DB 32 ETHPVGRVLVSIISIGPSPIEVYVRNQ 57

Search completed: July 16, 2003, 07:56:50
 Job time : 4.79677 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 0.910689 Seconds
(without alignments)
1594.036 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEVHPYGLPVGPGEPFKTLRVONLGCYVSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	1167	1 ITNG_HUMAN	075578 homo sapien
2	65.5	34.1	301	1 NODP_AZOBK	P28603 azospirillum
3	59	30.7	1189	1 ITNH_HUMAN	0944x5 homo sapien
4	54.5	28.4	299	1 NODP_RHISB	007308 rhizobium s
5	54	28.1	2148	1 VIRI_ABDAB	016927 aedes aegypt
6	52	27.1	101	1 ATPK_YEAST	006405 saccharomyc
7	52	27.1	348	1 SKI_MOUSE	006098 mus musculu
8	52	27.1	437	1 SKI_AVIES	P17863 avian eryth
9	52	27.1	728	1 SKI_HUMAN	P12755 homo sapien
10	52	27.1	750	1 SKI_CHICK	P49140 gallus gall
11	52	27.1	1111	1 GLI1_MOUSE	P47806 mus musculu
12	51.5	26.8	507	1 CP5G_CANTR	P30609 candida tro
13	51	26.6	402	1 YD29_SCHPO	013716 schizosacch
14	51	26.6	1053	1 SP5_SOUTU	043845 solanum tub
15	51	26.6	1064	1 NPEC2_MOUSE	Q06091 mus musculu
16	50	26.0	489	1 YS45_CAELI	009660 caenorhabdi
17	50	26.0	738	1 PLO3_HUMAN	006058 homo sapien
18	49.5	25.8	815	1 GYRB_MYXXA	033367 myxococcus
19	49	25.5	753	1 PMT3_YEAST	P47190 saccharomyc
20	49	25.5	1109	1 MSH7_ARATH	Q958v7 arabidopsi
21	48.5	25.3	352	1 ARGC_MYCTU	P94987 mycobacteri
22	48.5	25.3	653	1 YDE4_MYCTU	011034 mycobacteri
23	48.5	25.3	759	1 PMT6_YEAST	P42934 saccharomyc
24	48	25.0	387	1 TGT_MOUSE	Q91262 mus musculu
25	48	25.0	661	1 AIT2_VZVD	P03264 varicella-z
26	48	25.0	728	1 PLO1_RAT	Q63321 rattus norv
27	48	25.0	1289	1 CSAB_BACUD	Q4553 bacillus th
28	47.5	24.7	299	1 NODP_RHIME	P13441 rhizobium m
29	47.5	24.7	338	1 TRPD_RHOSH	Q9248 rhodospacer
30	47.5	24.7	625	1 TROR_MOUSE	Q08851 mus musculu
31	47.5	24.7	1065	1 KOGI_HUMAN	075912 homo sapien
32	47	24.5	434	1 N075_LUPLU	Q06841 lupinus lut
33	47	24.5	553	1 YF61_METUA	Q58956 methanococc

ALIGNMENTS

RESULT 1	ID	ITAG_HUMAN	STANDARD;	PRT;	1167 AA.
AC	075578; Q9UH28;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-0UN-2002 (Rel. 41, Last annotation update)				
DE	Integrin alpha-10 precursor.				
GN	ITGA10.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Articular chondrocytes;				
RX	MEDLINE=98352078; PubMed=9685391;				
RA	Casper L., Hellman U., Lundgren-Akerlund E.,				
RT	"Isolation, cloning, and sequence analysis of the integrin subunit				
RT	alpha10, a beta1-associated collagen binding integrin expressed on				
RT	chondrocytes."				
RL	J. Biol. Chem. 273:20383-20389 (1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Heart, and Endothelial cells;				
RX	MEDLINE=20169197; PubMed=10702680;				
RA	Lehner K., Ni J., Leung E., Gough S.M., Morris C.M., Lau D.,				
RT	Wang S.-X., Langley R., Krissansen G.W.;				
RT	"The integrin alpha10 subunit: expression pattern, partial gene				
RT	structure, and chromosomal localization."				
RL	Cytogenet. Cell Genet. 87:238-244 (1999).				
CC	- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.				
CC	- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10				
CC	ASSOCIATES WITH BETA-1.				
CC	- SUBCELLULAR LOCATION: Type I membrane protein.				
CC	- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN				
CC	MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.				
CC	- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWF DOMAIN. INTEGRINS				
CC	WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.				
CC	- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 VWF DOMAIN.				
CC	- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcements				
CC	or send an email to license@ebi.ac.uk).				
CC	-----				
DR	EMBL; AF074015; AAC1952.1; -				
DR	EMBL; AF112345; AAF21944.1; -				
DR	EMBL; AF127273; AAF61638.1; -				
DR	HSSP; P17301; IAOX.				
DR	GeneW; HGNC:6135; ITGA10.				

```

DR MIM; 604042; -.
DR InterPro: IPR000413; Integrin_alpha.
DR InterPro: IPR002035; VWF_A.
DR Pfam; PF00092; VWA; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int alpha; 5.
DR SMART; SM00327; VWA_1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS0234; VWF_A; 1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium.
FT SIGNAL 1 22
FT CHAIN 23 1167
FT DOMAIN 23 1122 INTEGRIN_ALPHA-10.
FT TRANSMEM 1123 1145 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1146 1167 POTENTIAL.
FT REPEAT 38 97 CYTOPLASMIC (POTENTIAL).
FT REPEAT 38 97 FG-GAP 1.
FT REPEAT 38 97 FG-GAP 2.
FT DOMAIN 167 350 VWF_A.
FT REPEAT 365 427 FG-GAP 3.
FT REPEAT 428 482 FG-GAP 4.
FT REPEAT 483 545 FG-GAP 5.
FT REPEAT 546 605 FG-GAP 6.
FT REPEAT 608 660 FG-GAP 7.
FT DOMAIN 1134 1140 POLY-LEU.
FT CA_BIND 494 502 POTENTIAL.
FT CA_BIND 558 566 POTENTIAL.
FT CA_BIND 620 628 POTENTIAL.
FT DISULFID 76 86 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 681 736 BY SIMILARITY.
FT DISULFID 789 795 BY SIMILARITY.
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 336 336 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 364 364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 733 733 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 763 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1011 1011 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1018 1018 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 844 844 I -> L (IN REF. 2).
FT CONFLICT 909 909 G -> V (IN REF. 2).
FT CONFLICT 926 926 E -> D (IN REF. 2).
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25CIAEAO CRC64;

Query Match 100.0%; Score 192; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 5.1e-19;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEVHPYGLPVGPGPEFKTLRVQNLGCVVSGLI 35
DB 952 YEVHPYGLPVGPGPEFKTLRVQNLGCVVSGLI 986

RESULT 2
NODP_AZOBR STANDARD; PRT; 301 AA.
AC P28603;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylyltransferase) (SAT) (ATP-sulfurylase small subunit) (Nodulation protein p).
GN NODP.
OS Azospirillum brasilense.
OC Plasmid.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;

```

```

OC Azospirillum.
OX NCBI_TaxID=192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sp7;
RC MEDLINE=92033082; PubMed=2131098;
RA Vielle C., Eimerich C.;
RT "Characterization of two Azospirillum brasilense Sp7 plasmid genes homologous to Rhizobium meliloti nodD2";
RL Mol. Plant Microbe Interact. 3:389-400(1990).
CC -!- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER TO NOD FACTOR.
CC -!- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
CC -!- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODQ, MAY BE PHYSICALLY ASSOCIATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
DR EMBL; M94886; AAA22185.1; -.
DR InterPro: IPR002500; PAPS_reduct.
DR Pfam; PF01507; PAPS_reduct.1.
KW Plasmid; Nodulation; transferase; Nucleotidyltransferase.
SQ SEQUENCE 301 AA; 34137 MW; 88D906E7B0237225 CRC64;

Query Match 34.1%; Score 65.5; DB 1; Length 301;
Best Local Similarity 48.4%; Pred. No. 0.075;
Matches 15; Conservative 5; Mismatches 8; Indels 3; Gaps 2;

QY 7 GTPVPGP-DEPKTLRVQNLGCVVSGLI 35
DB 228 GRLPLNPETPEMR-RVRFRLGCVPLSGAI 257

RESULT 3
ITAH_HUMAN STANDARD; PRT; 1189 AA.
AC G9UXK5; OSUKO1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-11 precursor.
GN ITGA11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal heart, and Osteoblast;
RC MEDLINE=99395147; PubMed=10464311;
RA Veiling T., Kusche-Gulberg M., Sejersten T., Gulberg D.;
RA Lehnert K., Ni J., Leung E., Gough S.M., Weaver A., Yao W.P., Liu D., Wang S.-X., Morris C.W., Kristiansen G.W.;
RT "Cloning, sequence analysis, and chromosomal localization of the novel human integrin alpha11 subunit (ITGA11).";
RL Genomics 60:179-187(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal muscle, and Uterus;
RC MEDLINE=99395147; PubMed=10464311;
RA Veiling T., Kusche-Gulberg M., Sejersten T., Gulberg D.;
RT "cDNA Cloning and Chromosomal Localization of Human alpha(11) Integrin. A collagen-binding, i domain-containing, beta(1)-associated integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
RN [3]
RP SEQUENCE OF 954-1188 FROM N.A.

```

RC TISSUE=Fibroblast; Andren N., Escivill X., Escarceller M., Sunoy L.;
 RA Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
 CC HEART, INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
 CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
 CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS, IN
 CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
 CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
 CC -1- DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
 CC FETAL MUSCLE CELLS (IN VITRO).
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
 CC WITH I-DOMAIN DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AF109681; AAF01258.1; -;
 CC EMBL, AF137378; AAD51919.2; -;
 CC EMBL, AL359064; CAB94392.1; -;
 CC HSSP; P17301; IAOX.
 CC Genew; HGNC:6136; ITGAL1.
 CC MIM; 604789; -;
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR002035; VFMA_A.
 CC Pfam; PF00092; vwa; 1.
 CC Pfam; PF01839; FG-GAP; 5.
 CC PRINTS; PR01185; INTEGRIN.
 CC PRINTS; PR00453; VFMADOMAIN.
 CC SMART; SM00191; Int_alpha; 5.
 CC SMART; SM00327; vwa; 1.
 CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
 CC PROSITE; PS50234; VFMA; 1.
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 CC Signal; Repeat; Calcium; Magnesium; Polymorphism.
 CC SIGNAL 1 22
 CC FT CHAIN 23 1189 INTEGRIN_ALPHA-11.
 CC FT DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 1143 1165 POTENTIAL.
 CC FT DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).
 CC REPEAT 38 94 FG-GAP 1.
 CC REPEAT 102 163 FG-GAP 2.
 CC REPEAT 167 345 VFMA.
 CC REPEAT 359 420 FG-GAP 3.
 CC REPEAT 422 475 FG-GAP 4.
 CC REPEAT 477 537 FG-GAP 5.
 CC REPEAT 539 598 FG-GAP 6.
 CC REPEAT 601 653 FG-GAP 7.
 CC DOMAIN 1154 1162 POLY-LEU.
 CC DOMAIN 1174 1177 POLY-ARG.
 CC CA_BIND 488 496 POTENTIAL.
 CC CA_BIND 551 559 POTENTIAL.
 CC CA_BIND 613 621 POTENTIAL.
 CC DISULFID 76 83 BY SIMILARITY.
 CC DISULFID 121 139 POTENTIAL.
 CC DISULFID 129 159 POTENTIAL.
 CC DISULFID 659 668 BY SIMILARITY.
 CC DISULFID 674 729 BY SIMILARITY.
 CC DISULFID 781 787 BY SIMILARITY.
 CC DISULFID 881 893 BY SIMILARITY.
 CC CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 95 95 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 291 311 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 331 331 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 642 642 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 694 694 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 884 894 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1032 1032 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 433 433 V -> M.
 FT VARIANT 524 524 R -> L.
 FT VARIANT 524 524 R -> L.
 FT VARIANT 972 972 L -> P.
 FT VARIANT 1003 1003 L -> P.
 FT VARIANT 1003 1003 I -> M.
 FT VARIANT 1030 1030 /FTID=VAR_009892.
 FT VARIANT 1030 1030 MISSING.
 FT VARIANT 1094 1094 /FTID=VAR_009893.
 FT VARIANT 1094 1094 L -> V.
 SQ SEQUENCE 1189 AA; 133609 MW; 60303C08A44C052 CRC64;
 Query Match 30.7%; Score 59; DB 1; Length 1189;
 Best Local Similarity 41.7%; Pred. No. 2.7;
 Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 Oy 12 GPGPEKFTTLRYONCGYVSGLI 35
 Db 981 GIGPEPSCIFRIQNLGLFPHGIM 1004
 RESULT 4
 ID NODE RHISB STANDARD; PRT; 299 AA.
 AC 007308;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4) (Sulfate adenylylate
 DE transferase) (SAT) (ATP-sulfurylase small subunit) (Modulation protein
 DE P).
 GN NODEP.
 OS Rhizobium sp. (strain BR816).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBT_Taxid:48291;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98083747; PubMed=9421916;
 RA Laereams T., Cooliaet N., Verreth C., Snoeck C., Hellings N.,
 RA Vanderleyden J., Martinez-Romero E.;
 RT "Functional redundancy of genes for sulphate activation enzymes in
 RT Rhizobium sp. BR816.";
 RL Microbiology 143:3933-3942(1997).
 CC -1- FUNCTION: PROPOSED TO PROVIDE ACTIVATED SULFATE FOR TRANSFER
 CC TO NOD FACTOR.
 CC -1- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.
 CC -1- SUBUNIT: SULFATE-ACTIVATING ENZYMES, NODP AND NODP, MAY BE
 CC PHYSICALLY ASSOCIATED (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSD SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@sb-sib.ch).
CC -----
CC DR EMBL: U59507; AAB95248.1;
CC DR InterPro: IPR002500; PAPS_reduct.
CC DR Pfam: PF01507; PAPS_reduct; 1.
CC KW Nucleotide transferase; Nucleotide transferase.
CC SQ SEQUENCE 299 AA; 34494 MW; 4F57C5F1983168F2 CRC64;

Query Match
Best Local Similarity 28.4%; Score 54.5; DB 1; Length 229;
Matches 11; Conservative 6; Mismatches 10; Indels 1; Gaps 1;

QY 9 LPVGPGE-FKTLRVQNLGCVVSGLI 35
Db 228 MPTEGEEVMDRVRFRITLGCYPLTGAI 255

RESULT 5
VITL AEDAE STANDARD; PRT; 2148 AA.
AC Q16927; Q1692;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vitellinogen A1 precursor (VG) (PVG1) [Contains: Vitellin light chain
DE (VL); Vitellin heavy chain (VH)].
GN VGAI.
OS Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Aedes.
OX NCBI_Taxid=7159;
[1]
RN RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=96035842; PubMed=7550249;
RA Romans P., Tu Z., Ke Z., Hagedorn H.H.;
RA "Analysis of a vitellinogen gene of the mosquito, Aedes aegypti and
RA comparisons to vitellinogenins from other organisms.";
RL Insect Biochem. Mol. Biol. 25:939-958(1995).
RN [2]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 17-24 AND 469-477.
RX STRAIN=UGALS; TISSUE=Fat body;
RX MEDLINE=94210487; PubMed=8158643;
RA Chen J.-S., Cho W.-L., Raikhel A.S.;
RA "Analysis of mosquito vitellinogenin cDNA. Similarity with vertebrate
RA phosphatins and arthropod serum proteins.";
RL J. Mol. Biol. 237:641-647(1994).
RN [3]
RN RP SULFATION.
RX TISSUE=Fat body;
RX MEDLINE=90277688; PubMed=2351682;
RA Dhadiatia T.S., Raikhel A.S.;
RA "Biosynthesis of mosquito vitellinogenin.";
RL J. Biol. Chem. 265:9924-9933(1990).
CC -1- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT. MAY SUPPLY AROMATIC AMINO
CC ACIDS TO THE CUTICLE OF RAPIDLY DEVELOPING EMBRYOS.
CC -1- TISSUE SPECIFICITY: PRODUCED BY THE FAT BODY, WHERE IT IS CLEAVED
CC IN THE ROUGH ENDOPLASMIC RETICULUM OR CIS-GOLGI BEFORE BEING
CC SECRETED INTO HEMOLYMPH. IT IS THEN SEQUESTERED BY A SINGLE CLASS
CC OF RECEPTOR MEDIATED ENDOCYTOSIS IN THE OVARY.
CC -1- INDUCTION: SYNTHESIZED ONLY BY SEXUALLY MATURE FEMALE AFTER
CC INGESTION OF BLOOD.
CC -1- PTM: GLYCOSYLATED. PHOSPHORYLATED AND SULFATED. THE LARGE SUBUNIT
CC IS SULFATED MORE EXTENSIVELY THAN THE SMALL ONE.
CC -1- POLYMORPHISM: ALLELIC VARIATIONS DETECTED IN THE MOSQUITO
CC POPULATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC DR EMBL: L41842; AAA99486.1;
CC DR EMBL: U02548; AAA18221.1;
CC DR InterPro: IPR001747; Lipid_transprt_N.
CC DR InterPro: IPR001846; VWF_D.
CC DR Pfam: PF000094; vwd; 1.
CC DR Pfam: PF01347; Vitellinogenin_N; 1.
CC DR SMART: SM00216; VWD; 1.
CC KW Glycoprotein; Phosphorylation; Storage protein; Signal; Polymorphism;
KW Sulfation.
FT SIGNAL 1 16
FT CHAIN 17 2148 VITELLIN LIGHT CHAIN.
FT CHAIN 17 468 VITELLIN HEAVY CHAIN.
FT CHAIN 469 2148 VITELLIN HEAVY CHAIN.
FT MOD_RES 138 138 SULFATION (POTENTIAL).
FT MOD_RES 142 142 SULFATION (POTENTIAL).
FT MOD_RES 1046 1046 SULFATION (POTENTIAL).
FT MOD_RES 1049 1049 SULFATION (POTENTIAL).
FT MOD_RES 1053 1053 SULFATION (POTENTIAL).
FT MOD_RES 1542 1542 SULFATION (POTENTIAL).
FT MOD_RES 1543 1543 SULFATION (POTENTIAL).
FT MOD_RES 1549 1549 SULFATION (POTENTIAL).
FT MOD_RES 1716 1716 SULFATION (POTENTIAL).
FT MOD_RES 1785 1785 SULFATION (POTENTIAL).
FT MOD_RES 1788 1788 SULFATION (POTENTIAL).
FT MOD_RES 1801 1801 SULFATION (POTENTIAL).
FT MOD_RES 1803 1803 SULFATION (POTENTIAL).
FT MOD_RES 1867 1867 SULFATION (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1119 1119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1315 1315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1631 1631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1956 1956 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 417 440 POLY-SER.
FT DOMAIN 496 512 POLY-SER.
FT DOMAIN 520 548 POLY-SER.
FT DOMAIN 2007 2025 POLY-SER.
FT DOMAIN 2033 2042 POLY-SER.
FT CONFLICT 39 47 MISSING (IN REF. 2).
FT CONFLICT 906 906 R -> S (IN REF. 2).
FT CONFLICT 1367 1367 F -> Y (IN REF. 2).
FT CONFLICT 1576 1576 C -> Y (IN REF. 2).
FT CONFLICT 1757 1757 N -> Y (IN REF. 2).
FT CONFLICT 1935 1935 A -> P (IN REF. 2).
FT CONFLICT 2031 2031 G -> S (IN REF. 2).
SQ SEQUENCE 2148 AA; 250249 MW; 2AB4DF63AC766AF CRC64;

Query Match
Best Local Similarity 28.1%; Score 54; DB 1; Length 2148;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 EVHPGTLVPVGPGEF 18
Db 375 DVHPYDKVMPGPAADLK 391

RESULT 6
ATPK_YEAST STANDARD; PRT; 101 AA.
AC Q06405;

```



```

RL Mol. Cell. Biol. 9:4038-4045(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28490; AAA9142.1; -
DR PIR; A32574; TVFSK.
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
KW Oncogene.
SQ SEQUENCE 437 AA; 48867 MW; A29372928E0036F8 CRC64;

Query Match
Best Local Similarity 35.3%; Score 52; DB 1; Length 437;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 46 PVMPGFPMPSDRSTERCETVLEGRTISCFVVG 79

RESULT 9
SKI_HUMAN STANDARD; PRT; 728 AA.
AC P12755;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ski oncogene (C-ski).
GN SKI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345144; PubMed=2762147;
RA Nomura N., Sasamoto S., Ishii S., Matsui M., Ishizaki R.;
RT "Isolation of human cDNA clones of ski and the ski-related gene,
RT sno.";
RL Nucleic Acids Res. 17:5489-5500(1989).
CC - FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL
CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC
CC LINEAGE.
CC - SUBUNIT: INTERACTS WITH SMAD2, 3 AND 4.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SMO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15218; CAA33288.1; -
DR PIR; S06053; TVHUSK.
DR TRANSFAC; T04643; -
DR Genew; HGNC:10896; SKI.
DR MIM; 164780; -
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
KW Proto-oncogene; Coiled coil; Nuclear protein; Repeat.
FT DOMAIN 536 710
SQ SEQUENCE 728 AA; 80004 MW; 9B78C4940A28C2DA CRC64;

```

```

Query Match
Best Local Similarity 35.3%; Score 52; DB 1; Length 728;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 84 PVMPGFPMPSDRSTERCETVLEGRTISCFVVG 117

RESULT 10
SKI_CHICK STANDARD; PRT; 750 AA.
ID SKI_CHICK
AC P49140;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ski oncogene (C-ski).
GN SKI.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89384635; PubMed=2779576;
RA Suttrave P., Hughes S.H.;
RT "Isolation and characterization of three distinct cDNAs for the
RT chicken c-ski gene.";
RL Mol. Cell. Biol. 9:4046-4051(1989).
CC - FUNCTION: MAY PLAY A ROLE IN TERMINAL DIFFERENTIATION OF SKELETAL
CC MUSCLE CELLS BUT NOT IN THE DETERMINATION OF CELLS TO THE MYOGENIC
CC LINEAGE.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SMO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28517; AAA8730.1; -
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
KW Proto-oncogene; Coiled coil; Nuclear protein; Repeat.
FT DOMAIN 558 732
SQ SEQUENCE 750 AA; 84282 MW; 01E322EBE0C402B CRC64;

Query Match
Best Local Similarity 35.3%; Score 52; DB 1; Length 750;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVONLCGYVSG 33
Db 67 PVMPGFPMPSDRSTERCETVLEGRTISCFVVG 100

RESULT 11
GLI_MOUSE STANDARD; PRT; 1111 AA.
ID GLI_MOUSE
AC P47806; O8QYK;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein Gli1 (Glioma-associated oncogene homolog).
GN GLI OR GLI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

CC NCBI_TaxID=10090;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC MEDLINE=98192509; PubMed=9524201;
 CC Liu C.Z., Yang J.T., Yoon J.W., Walterhouse D., Iannaccone P.;
 CC "Characterization of the promoter region and genomic organization of
 CC GLI1, a member of the Sonic hedgehog-Patched signaling pathway.";
 CC Gene 209:1-11(1998).
 CC (2)
 CC SEQUENCE FROM N.A.
 CC STRAIN=ICR;
 CC MEDLINE=99365280; PubMed=10433919;
 CC Saeki H., Nishizaki Y., Hui C., Nakafuku M., Kondoh H.;
 CC "Regulation of Gli2 and Gli3 activities by an amino-terminal
 CC repression domain: implication of Gli2 and Gli3 as primary mediators
 CC of Shh signaling.";
 CC Development 126:3915-3924(1999).
 CC (3)
 CC SEQUENCE OF 272-837 FROM N.A.
 CC STRAIN=CD-1; TISSUE=Embryo;
 CC MEDLINE=99337281; PubMed=8364225;
 CC Waterhouse D., Ahmed M., Slusarski D., Kalamaras J., Boucher D.,
 CC Holmgren R., Iannaccone P.;
 CC "Gli, a zinc finger transcription factor and oncogene, is expressed
 CC during normal mouse development.";
 CC Dev. Dyn. 196:91-102(1993).
 CC -1- FUNCTION: MAY REGULATE THE TRANSCRIPTION OF SPECIFIC GENES DURING
 CC NORMAL DEVELOPMENT. MAY PLAY A ROLE IN CRANIOFACIAL DEVELOPMENT
 CC AND DIGITAL DEVELOPMENT, AS WELL AS DEVELOPMENT OF THE CENTRAL
 CC NERVOUS SYSTEM AND GASTROINTESTINAL TRACT. IMPLICATED IN THE
 CC TRANSDUCTION OF SHH SIGNAL.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: UTERUS, BRAIN, TESTIS AND LIMB.
 CC -1- DEVELOPMENTAL STAGE: IS DETECTED ON DAYS 10 THROUGH 18 OF
 CC EMBRYONIC DEVELOPMENT. DURING GESTATION IT IS DETECTED IN MECKELS
 CC PRECAECAL MESENCHYME, THE BASIS OCCIPITUS, RIB MESENCHYMAL
 CC CONDENSATIONS, PRIMORDIAL VERTEBRAL BODIES, DIGITAL MESENCHYMAL
 CC CONDENSATIONS IN FOREFOOT AND HINDFOOT PLATES, THE EPENDYMAL LAYER
 CC OF THE SPINAL CORD, AND THE MESODERM OF THE GASTROINTESTINAL
 CC TRACT. EXPRESSION PERSISTS THROUGHOUT GESTATION IN DEVELOPING BONE
 CC AND CARTILAGE OF THE EXTREMITIES, THE RIBS, AND THE VERTEBRAL
 CC BODIES AS WELL AS THE GASTROINTESTINAL TRACT MESODERM.
 CC -1- SIMILARITY: BELONGS TO THE GLI FAMILY OF C2H2-TYPE ZINC-FINGER
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF026305; AAC09169.1; -;
 CC EMBL; AB025922; BAA85004.1; -;
 CC HSSP; P08151; 2GLI.
 CC MGD; MGI:95727; Gli1.
 CC InterPro: IPR000822; Znf_C2H2.
 CC Pfam: PF00096; ZF-C2H2; 5.
 CC SMART; SM00355; ZNF_C2H2; 5.
 CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
 CC K1 zinc-finger; Metal-binding; DNA-binding; Transcription regulation;
 CC Nuclear protein.
 CC DOMAIN 228 390 ZINC FINGERS.
 CC ZN_FING 228 263 C2H2-TYPE.
 CC ZN_FING 271 298 C2H2-TYPE.
 CC ZN_FING 304 328 C2H2-TYPE.
 CC ZN_FING 334 359 C2H2-TYPE.
 CC ZN_FING 365 390 C2H2-TYPE.
 CC DOMAIN 1042 1059 ASP/GLU-RICH (ACIDIC).
 CC CONFLICT 154 154 V -> L (IN REF. 1).
 CC CONFLICT 170 170 H -> Y (IN REF. 1).
 CC FT

FT CONFLICT 173 173 A -> S (IN REF. 1).
 FT CONFLICT 179 179 T -> I (IN REF. 1).
 FT CONFLICT 194 194 P -> R (IN REF. 1).
 FT CONFLICT 210 210 I -> T (IN REF. 1).
 FT CONFLICT 271 271 F -> S (IN REF. 1).
 FT CONFLICT 475 475 MISSING (IN REF. 1 AND 3).
 FT CONFLICT 567 569 FPP -> LPT (IN REF. 1 AND 3).
 FT CONFLICT 707 707 E -> D (IN REF. 1 AND 3).
 FT CONFLICT 777 777 MISSING (IN REF. 1 AND 3).
 FT CONFLICT 837 837 H -> P (IN REF. 1 AND 3).
 FT CONFLICT 864 864 G -> V (IN REF. 1).
 FT CONFLICT 919 920 GU -> RA (IN REF. 1).
 FT CONFLICT 936 936 S -> Y (IN REF. 1).
 FT CONFLICT 950 952 AAA -> RR (IN REF. 1).
 FT CONFLICT 967 967 G -> R (IN REF. 1).
 FT CONFLICT 990 990 P -> A (IN REF. 1).
 FT CONFLICT 1029 1029 A -> P (IN REF. 1).
 FT CONFLICT 1062 1063 OG -> R (IN REF. 1).
 FT CONFLICT 1111 AA; 118555 MW; 0DBAB163CAF27289 CRC64;
 FT SEQUENCE
 SQ
 Query Match 27.1%; Score 52; DB 1; Length 1111;
 Best Local Similarity 60.0%; Pred. No. 24;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 YEHPYGLPYGPGP 15
 DB 744 YEARGPGSLPQGP 758
 RESULT 12
 ID CP5G CANTR STANDARD; PRT; 507 AA.
 AC P30609;
 DT 01-APR-1993 (Rel. 25. Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytochrome P450 52A7 (EC 1.14.14.-) (CYPL1A7) (Alkane-inducible P450-
 DE ALK4).
 GN CYP52A7.
 OS Candida tropicalis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=54482;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 750;
 RX MEDLINE=93090274; PubMed=1457045;
 RA Seghezzi W., Melli C., Ruffiner R., Kuenzi R., Sanglard D.,
 RA Fiechter A.;
 RT "Identification and characterization of additional members of the
 RT cytochrome P450 multigene family CYP52 of Candida tropicalis.";
 RL DNA Cell Biol. 11:767-780(1992).
 CC -1- FUNCTION: TOGETHER WITH AN NADPH CYTOCHROME P450 THE ENZYME SYSTEM
 CC CATALYZES THE TERMINAL HYDROXYLATION AS THE FIRST STEP IN THE
 CC ASSIMILATION OF ALKANES AND FATY ACIDS. PREFERENTIALLY
 CC HYDROXYLATES LAURIC ACID.
 CC -1- CATALYTIC ACTIVITY: Hydroxylation of N-alkanes at the terminal
 CC position.
 CC -1- INDUCTION: BY VARIOUS ALKANES.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z13011; CAA78355.1; -;
 CC PIR; S22973; S22973.
 CC HSSP; P14779; 10PZ.
 CC InterPro: IPR001128; Cytochrome_P450.
 CC DR

DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME P450; 1.
 KW Electron transport; Oxidoreductase; Monooxygenase; Heme;
 KM Transmembrane.
 FT TRANSMEM 6 26 POTENTIAL.
 FT BINDING 456 456 HEME (BY SIMILARITY).
 SQ SEQUENCE 507 AA; 58671 MW; 34B3DDE3D1E5C87 CRC64;

Query Match 26.8%; Score 51.5; DB 1; Length 507;
 Best Local Similarity 40.7%; Pred. No. 12;
 Matches 11; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

QY 8 TLPGGPEFKTLRVONLCGYV-VSG 33
 DB 388 TLPGGPGDKOPILVRKMSCSIFISG 414

RESULT 13
 YD29 SCHPO STANDARD; PRT; 402 AA.
 AC 013716;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein C14C4.09 in chromosome 1.
 GN SPAC14C4.09.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NC NCB1_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambut R., Furnelle B.,
 RA Goffeau A., Cadieu E., Diano S., Gloux S., Lelure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Banito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpkovski G.V., Useery D., Barrett B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z89596; CAB11202.1;
 DR InterPro: IPR005197; Glyco_hydro_71.

DR Pfam: PF03659; Glyco_hydro_71; 1.
 KW Hypothetical protein.
 KW SEQUENCE 402 AA; 44515 MW; DCB13CDB5E070B0 CRC64;
 SQ SEQUENCE

Query Match 26.6%; Score 51; DB 1; Length 402;
 Best Local Similarity 43.5%; Pred. No. 11;
 Matches 10; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

QY 7 GTLPVGPPEFKTLRVONLCGY 29
 DB 374 GTLPVGPPEFKTLRVONLCGY 396

RESULT 14
 SPS SOLTU STANDARD; PRT; 1053 AA.
 ID SPS SOLTU
 AC 048545;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Sucrose-phosphate synthase (EC 2.4.1.14) (UDP-glucose-fructose-
 DE phosphate glucosyltransferase).
 GN SPS.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
 NC NCB1_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Deslee; TISSUE=leaf;
 RX MEDLINE=95201832; PubMed=7894514;
 RA Zrenner R., Salanoubat M., Willmitzer L., Somerville U.,
 RT "Evidence of the crucial role of sucrose synthase for sink strength
 RT using transgenic potato plants (Solanum tuberosum L.).",
 RL Plant J. 7:97-107(1995).

CC -1- FUNCTION: INVOLVED IN THE REGULATION OF CARBON PARTITIONING IN
 CC THE LEAVES OF PLANTS. MAY REGULATE THE SYNTHESIS OF SUCROSE AND
 CC THEREFORE PLAY A MAJOR ROLE AS A LIMITING FACTOR IN THE EXPORT OF
 CC PHOTOSYNTHETIC PRODUCTS OUT OF THE LEAF.
 CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-fructose 6-phosphate = UDP +
 CC sucrose 6-phosphate.
 CC -1- ENZYME REGULATION: ACTIVITY REGULATED BY PHOSPHORYLATION AND
 CC MODERATED BY CONCENTRATION OF METABOLITES AND LIGHT.
 CC -1- PATHWAY: Sucrose synthesis.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER (BY SIMILARITY).
 CC -1- PFM: PHOSPHORYLATED. HOWEVER, PHOSPHORYLATION IS NOT ESSENTIAL FOR
 CC ENZYME FUNCTION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X73477; CAAS1872.1;
 DR InterPro: IPR01296; Glycosyltransf_1.
 DR Pfam: PF00534; Glycosyltransf_1;
 KW Transferase; Glycosyltransferase; Phosphorylation.
 SQ SEQUENCE 1053 AA; 118292 MW; D6C933798567820A CRC64;

Query Match 26.6%; Score 51; DB 1; Length 1053;
 Best Local Similarity 35.5%; Pred. No. 31;
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEHPVGTLPVGPPEFKTLRVONLCGYV 31
 DB 911 FKVCKPGTVP--PSKELRKVRIOALRCHAV 939

ID	NCFC_MOUSE	STANDARD;	PRT;	1064 AA.
AC	Q60591	Q60985; Q60984;		
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Nuclear factor of activated T-cells, cytoplasmic 2 (T cell transcription factor NFAT1) (NFAT pre-existing subunit) (NF-ATp).			
GN	NFATC2 OR NFAT1 OR NFATP.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM A).			
RP	MEDLINE=94053710; PubMed=6235597;			
RX	McCaffrey P.G., Luo C., Keppolla T.K., Jain J., Badalian T.M., Ho A.M., Burgon E., Lane W.S., Lambert J.N., Curran T., Verdine G.L., Rao A., Hogan P.G.,			
RA	"Isolation of the cyclosporin-sensitive T cell transcription factor NFATP."			
RT	Science 262:750-754(1993).			
RL	[2] SEQUENCE FROM N.A. (ISOFORMS A; B AND C).			
RN	MEDLINE=96251346; PubMed=8668213;			
RX	Luo C., Burgon E., Carew J.A., McCaffrey P.G., Badalian T.M., Lane W.S., Hogan P.G., Rao A.,			
RA	"Recombinant NFAT1 (NFATP) is regulated by calcineurin in T cells and mediates transcription of several cytokine genes."			
RT	Mol. Cell. Biol. 16:3955-3966(1996).			
RL	[3] MUTAGENESIS OF ARG-423, HIS-425, TYR-426, THR-428 AND GLU-429. MEDLINE=95181386; PubMed=7876165;			
RN	Jain J., Burgon E., Badalian T.M., Hogan P.G., Rao A.;			
RA	"A similar DNA-binding motif in NFAT family proteins and the Rel homology region."			
RT	J. Biol. Chem. 270:4138-4145(1995).			
RL	[4] MUTAGENESIS OF ARG-112; GLU-114 AND THR-116. MEDLINE=98325377; PubMed=9660947;			
RN	Atamburu J., Garcia-Cozar F., Raghavan A., Okamura H., Rao A., Hogan P.G.;			
RA	"Selective inhibition of NFAT activation by a peptide spanning the calcineurin targeting site of NFAT."			
RT	Mol. Cell 1:627-637(1998).			
RL	[5] REVIEW			
RP	MEDLINE=99189746; PubMed=10089876;			
RX	Ciebtrey G.R.;			
RA	"Genetic signals and specific outcomes: signaling through Ca2+, calcineurin, and NF-AT."			
RT	Cell 96:611-614(1999).			
RL	[6] FUNCTION. PLAYS A ROLE IN THE INDUCIBLE EXPRESSION OF CYTOKINE GENES IN T CELLS, ESPECIALLY IN THE INDUCTION OF THE IL-2, IL-3, IL-4, TNF-ALPHA OR GM-CSF.			
CC	SUBUNIT. MEMBER OF THE MULTICOMPONENT NFATC TRANSCRIPTION COMPLEX THAT CONSISTS OF AT LEAST TWO COMPONENTS, A PRE-EXISTING CYTOPLASMIC COMPONENT NFATC3 AND AN INDUCIBLE NUCLEAR COMPONENT NFATC1. OTHER MEMBERS SUCH AS NFATC4, NFATC3 OR MEMBERS OF THE ACTIVATING PROTEIN-1 FAMILY, MAP, GATA4 AND C/EB-300 CAN ALSO BIND TO THE COMPLEX. NFATC PROTEINS BIND TO DNA AS MONOMERS.			
CC	SUBCELLULAR LOCATION: CYTOPLASMIC FOR THE PHOSPHORYLATED FORM AND NUCLEAR AFTER ACTIVATION THAT IS CONTROLLED BY CALCINEURIN-MEDIATED DEPHOSPHORYLATION. RAPID NUCLEAR EXIT OF NFATC IS THOUGHT TO BE ONE MECHANISM BY WHICH CELLS DISTINGUISH BETWEEN SUSTAINED AND TRANSIENT CALCIUM SIGNALS. THE SUBCELLULAR LOCALIZATION OF NFATC PLAY A KEY ROLE IN THE GENE TRANSCRIPTION.			
CC	ALTERNATIVE PRODUCTS: 3 ISOFORMS; A (SHOWN HERE), B AND C; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	TISSUE SPECIFICITY: EXPRESSED IN THYMUS, SPLEEN, HEART, TESTIS, BRAIN, PLACENTA, MUSCLE AND PANCREAS.			
CC	DOMAIN: REL SIMILARITY DOMAIN (RSD) ALLOWS DNA-BINDING AND			

```
CC COOPERATIVE INTERACTIONS WITH APL FACTORS (BY SIMILARITY) .  
CC -I- PPM: PHOSPHORYLATED BY NEATC-KINASE; DEPHOSPHORYLATED BY  
CC CALCINEURIN (BY SIMILARITY)  
CC --I- SIMILARITY: BELONGS TO THE REL/DORSAAL FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outpost at -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/  
CC or send an email to license@isb.sib.ch).  
CC -----  
DR EMBL; U02079; AAC52929.1; -.  
DR EMBL; U36575; AAC52930.1; -.  
DR EMBL; U36575; AAC52931.1; -.  
MR MGD; MGJ:U02463; NFEKTC2.  
DR InterPro; IPR002909; IPT_TIG.  
DR InterPro; IPR000451; NF_Rel_dor_fam.  
DR Pfam; PF00554; RHD; 1.  
DR Pfam; PF01833; TIG; 1.  
DR SMART; SMO0429; IPT; 1.  
DR PROSITE; PS01204; REL_1; FALSE_NEG.  
KW PROSITE; PSS0254; REL_2; 1.  
KM Transcription regulation; Activator; Nuclear protein; DNA-binding;  
KW Alternative splicing; Phosphorylation; Repeat.  
FT DOMAIN 111 116  
FT FT DOMAIN 119 201  
FT TRANS-ACTIVATION DOMAIN A (TAD-A) .  
FT REPEAT 186 292  
FT REPEAT 215 202  
FT REPEAT 274 231  
FT REPEAT 274 230  
FT MOD_DOMAIN 423 430  
FT MOD_RES 110 110  
FT VARSPLIC 910 1064  
  
PHEOPHORYLATYON (BY SIMILARITY).  
AAATESVAVGTERIERFERFKKTLPQGLPSFLLGSISAG  
PRSCPTSEPKRIEDVDPLSGSQIAMCOHPGTCPVLGGPL  
AVEWMEQLGRGLEPIPAWPDSAGSLHEVSGLGVVGVM  
ALTITMHFSMDONTSPFHNRKHVEVASPGMI -> ELIDT  
HLSTAIQNLT (IN ISOFORM B).  
HAATESVAVGTERIERFERFKKTLPQGLPSFLLGSISAG  
PRSCPTSEPKRIEDVDPLSGSQIAMCOHPGTCPVLGGPL  
AVEWMEQLGRGLEPIPAWPDSAGSLHEVSGLGVVGVM  
ALTITMHFSMDONTSPFHNRKHVEVASPGMI -> VNEII  
RKESGPSRNOT (IN ISOFORM C).  
R->-A: LOWERS DEPHOSPHORYLATION.  
E->-A: LOWERS DEPHOSPHORYLATION.  
T->-A: NO DEPHOSPHORYLATION.  
R->-A: DECREASE IN-BINDING TO DNA.  
H->-A: NO CHANGE IN BINDING TO DNA.  
Y->-A: DECREASE IN BINDING TO DNA.  
T->-A: NO CHANGE IN BINDING TO DNA.  
T->-C: NO CHANGE IN BINDING TO DNA AND  
CONTERS DNA-BINDING SENSITIVITY TO  
SUPEHYDRL MODIFICATIONS.  
B->-A: DECREASE IN BINDING TO DNA.  
CLTSIC16DDID3CA6FO CRC64;
```

Query March Best Local Similarity 26.6%; Score 51; DB 1; Length 1064;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Dy 4 HPVTGLPVGRP 15
||| |||
Db 979 HFLGTCPVLGRP 990

Search completed: July 16, 2003, 07:52:51
Job time : 2.91069 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 / Search time 3.59353 Seconds
(without alignments)
2006.842 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHHPYGTLPVGPGEFFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	30.7	823	4 Q8WY18	Q8WY18 homo sapien
2	58.5	30.5	2275	12 Q8Q841	Q8Q841 chimpanzee
3	58	30.2	695	13 Q9Y102	Q9Y102 brachydanio
4	56	29.2	94	15 Q9Y101	Q9Y101 human immun
5	55	28.6	1054	10 Q9SNY7	Q9SNY7 nicotiana t
6	54.5	28.4	1173	17 Q8ZX19	Q8ZX19 pyrobaculum
7	54	28.1	699	5 O15816	O15816 dictyosteli
8	53	27.6	94	15 Q9Y121	Q9Y121 human immun
9	53	27.6	714	13 Q9P031	Q9P031 oreochromis
10	53	27.6	727	13 Q9Y103	Q9Y103 brachydanio
11	52	27.1	94	15 Q9Y125	Q9Y125 human immun
12	52	27.1	359	13 Q8UMC1	Q8UMC1 human immun
13	52	27.1	461	17 Q979H6	Q979H6 gallus gall
14	52	27.1	725	11 Q8V115	Q8V115 thermoplasma
15	52	27.1	730	6 Q9TUG2	Q9TUG2 mus musculus
16	52	27.1	730	6 Q9TUG2	Q9TUG2 equus caball

17	51.5	26.8	390	5 P91571	P91571 caenorhabdi
18	51	26.6	94	15 Q9Y134	Q9Y134 human immun
19	51	26.6	312	12 Q8UZD8	Q8UZD8 cercopithec
20	51	26.6	366	7 Q8SNF2	Q8SNF2 sigmodon hi
21	51	26.6	384	13 Q73714	Q73714 brachydanio
22	51	26.6	741	11 Q9CY19	Q9CY19 mus musculu
23	51	26.6	741	11 Q9ROE1	Q9ROE1 mus musculu
24	50.5	26.3	300	2 Q68031	Q68031 rhodobacter
25	50.5	26.3	317	16 Q9PD79	Q9PD79 xyloella fas
26	50	26.0	93	15 Q9Y153	Q9Y153 human immun
27	50	26.0	94	15 Q9Y155	Q9Y155 human immun
28	50	26.0	94	15 Q9Y146	Q9Y146 human immun
29	50	26.0	94	15 Q9Y145	Q9Y145 human immun
30	50	26.0	94	15 Q9Y140	Q9Y140 human immun
31	50	26.0	94	15 Q9Y139	Q9Y139 human immun
32	50	26.0	94	15 Q9Y137	Q9Y137 human immun
33	50	26.0	94	15 Q9Y131	Q9Y131 human immun
34	50	26.0	94	15 Q9Y126	Q9Y126 human immun
35	50	26.0	94	15 Q9Y153	Q9Y153 human immun
36	50	26.0	94	15 Q9Y155	Q9Y155 human immun
37	50	26.0	142	2 Q9AH33	Q9AH33 pseudomonas
38	50	26.0	142	2 Q9RIH7	Q9RIH7 pseudomonas
39	50	26.0	149	10 Q9FSN5	Q9FSN5 oryza sativ
40	50	26.0	191	17 Q8TVD9	Q8TVD9 methanopyru
41	49.5	25.8	94	15 Q9Y141	Q9Y141 human immun
42	49.5	25.8	348	17 Q97V15	Q97V15 sulfolobus
43	49.5	25.8	361	6 Q95J14	Q95J14 macaca fasc
44	49.5	25.8	398	5 Q8SVW6	Q8SVW6 encephalito
45	49.5	25.8	585	6 Q95LZ0	Q95LZ0 macaca fasc

ALIGNMENTS

RESULT 1	Q8WY18	PRELIMINARY;	PRT;	823 AA.
AC	Q8WY18;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	MSTP018.			
GN	MST018.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	TISSUE=AOETA;			
RA	Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,			
RA	Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.D.,			
RA	Gao R.L., Qiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,			
RA	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
RL	EMBL: AF11799; AAL39001.1; -			
DR	InterPro: IPR000413; Integrin_alpha.			
DR	Pfam: PF01839; FG-GAP; 4.			
DR	PRINTS: PRO1185; INTEGRIN.			
DR	SMART: SM00191; Int_alpha; 4.			
SQ	SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;			
Query Match	30.7%; Score 59; DB 4; Length 823;			
Best Local Similarity	41.7%; Pred. No. 7.4;			
Matches	10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;			
Qy	12 GPGEFFKTLRVQNLGCVVSGLI 35			
DB	616 GIGPFFCFRIQNLGLFPIHGM 639			
RESULT 2	Q8Q841			

ID Q8Q541 PRELIMINARY; PRT; 2275 AA.
 AC Q8Q541;
 DT 01-JUN-2002 (TRENBLREL. 21, Created)
 DT 01-JUN-2002 (TRENBLREL. 21, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Tegument protein UL48.
 OS Chimpazee cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OC NCBI_TaxID=188763;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,
 RA Alencor D.J., Hayward G.S., McGeoch D.J.;
 RA "The human cytomegalovirus genome revisited."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF480884; AF480697.1;
 SQ SEQUENCE 2275 AA; 255990 MW; 45BBA419CA576BCD CRC64;

Query Match 30.5%; Score 58.5; DB 12; Length 2275;
 Best Local Similarity 34.0%; Pred. No. 28;
 Matches 16; Conservative 7; Mismatches 7; Indels 17; Gaps 3;

Qy 1 YEVHPY-----GTLVPGPG--PEFKTLRVONTGCVV 31
 DB 1680 YDLKPYFSQQQQQRAQDGSVPLGPAPPEANTLNLK-LFCYV 1725

RESULT 3

Q9Y102 PRELIMINARY; PRT; 695 AA.
 ID Q9Y102;
 AC Q9Y102;
 DT 01-MAY-1999 (TRENBLREL. 10, Created)
 DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Nuclear oncoprotein.
 GN SKIB OR SKIB.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20368164; PubMed=10906458;
 RA Kaufman C.D., Martinez-Rodriguez G., Hackelt P.B. Jr.;
 RT "Ecopic expression of c-ski disrupts gastrulation and neural
 RT patterning in zebrafish."
 RL Mech. Dev. 95:147-162(2000).
 DR EMBL; AF060118; AAC64707.1;
 DR ZFIN; ZDB-GENE-990715-10; skib.
 DR InterPro: IPR003380; Transform_Ski.
 DR Pfam; PF02437; Ski_Sno; 1.
 SQ SEQUENCE 695 AA; 79135 MW; 9BD424C32F61C4ED CRC64;

Query Match 30.2%; Score 58; DB 13; Length 695;
 Best Local Similarity 35.9%; Pred. No. 8.5;
 Matches 14; Conservative 4; Mismatches 11; Indels 10; Gaps 1;

Qy 5 PYGTLPVGPGRPF-----KTLRVONTGCVVSG 33
 DB 58 PAQTPVMPGPFIPSDRSTERCETVLERETISCFVVG 96

RESULT 4

Q9YTS1 PRELIMINARY; PRT; 94 AA.
 ID Q9YTS1;
 AC Q9YTS1;
 DT 01-MAY-1999 (TRENBLREL. 10, Created)
 DT 01-MAY-1999 (TRENBLREL. 10, Last sequence update)
 DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
 DE Envelope-glycoprotein (Fragment).
 GN ENV.

OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_TaxID=11676;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S4V2-6;
 RX MEDLINE=98445411; PubMed=9770526;
 RA Markham R.B., Wang W.C., Weisstein A.E., Wang Z., Munoz A.,
 RA Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,
 RA Yu X.F.;
 RT "Patterns of HIV-1 evolution in individuals with differing rates of
 RT CD4 T cell decline."
 RT Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).
 DR EMBL; AF089156; AAC78866.1;
 DR InterPro: IPR000777; GP120.
 DR Pfam; PF00516; GP120.1.
 KM AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 94 AA; 10841 MW; 27E1BE1608E890C4 CRC64;

Query Match 29.2%; Score 56; DB 15; Length 94;
 Best Local Similarity 47.4%; Pred. No. 1.7;
 Matches 9; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 8 TLVPGPGPEFKTLRVONTL 26
 DB 38 SIPIGPGARFTTGRIKNI 56

RESULT 5

Q9SNV7 PRELIMINARY; PRT; 1054 AA.
 ID Q9SNV7;
 AC Q9SNV7;
 DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE Sucrose-6-phosphate synthase (EC 2.4.1.14).
 GN SPS.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OC NCBI_TaxID=4097;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CV. SAMSUN NN;
 RA Borker F.;
 RT "Cloning and Characterization of a Sucrose-6-phosphate Synthase from
 RT Nicotiana tabacum."
 RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF194032; AAF06792.1;
 DR InterPro: IPR001296; Glycosyltransf_1.
 DR Pfam; PF00534; Glycosyltransf_1.
 DR KMW Glycosyltransferase; Transfease.
 SQ SEQUENCE 1054 AA; 118692 MW; 2C593AF27649A26D CRC64;

Query Match 28.6%; Score 55; DB 10; Length 1054;
 Best Local Similarity 35.5%; Pred. No. 37;
 Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEVHPYGTLPVGPGRPF-----KTLRVONTGCVV 31
 DB 912 FKVRKLTGTF--PAKELRKLRIQALCHAV 940

RESULT 6

Q8ZXI9 PRELIMINARY; PRT; 1173 AA.
 ID Q8ZXI9;
 AC Q8ZXI9;
 DT 01-MAR-2002 (TRENBLREL. 20, Created)
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)

DE Molybdopterin oxidoreductase, molybdopterin binding subunit.
GN PAE1265.
OS Pyrobaculum aerophilum.
AC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxId=13773;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX PubMed=11792869;
RA Flier-Gibson S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
aerophilum.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
DR EMBL; AB009811; AAL6359.1; -.
DR InterPro; IPR000566; Lipocin_cytfABP.
DR InterPro; IPR001467; Prok_Mboxred.
DR Pfam; PF00384; molybdopterin; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 1173 AA; 132457 MW; 790F6CC747D512BC CRC64;

Query Match 28.4%; Score 54.5; DB 17; Length 1173;
Best Local Similarity 41.4%; Pred. No. 50;
Matches 12; Conservative 4; Mismatches 10; Indels 3; Gaps 1;

QY 10 PVGPGPEFKTLRV---QNLGCVVSGLI 35
DB 38 PYGDPQFGTNRVYVSSCLGCVRCGIV 66

RESULT 7
015816 PRELIMINARY; PRT; 699 AA.
AC 015816;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Myb2 protein.
GN MYB2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=98283920; PubMed=9620859;
RA Otsuka H., Van Haastert P.J.M.;
RT "A novel Myb homologue initiates Dictyostelium development by
RT induction of adenyl cyclase expression.";
RL Genes Dev. 12:1738-1748(1998).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 3 MYB-LIKE DOMAINS.
DR EMBL; AJ002383; CA05357.1; -.
DR HSSP; P06876; IMBG.
DR InterPro; IPR001005; Myb DNA-binding.
DR Pfam; PF00249; myb DNA-binding; 3.
DR SMART; SM00395; SANT; 3.
DR PROSITE; PS50090; MYB_3; 3.
DR DNA-binding; Nuclear protein.
SQ SEQUENCE 699 AA; 80381 MW; EA989DC77953F265 CRC64;

Query Match 28.1%; Score 54; DB 5; Length 699;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 12; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

RESULT 8
09YT21 PRELIMINARY; PRT; 94 AA.
ID 09YT21
AC 09YT21;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxId=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S4V4-10;
RX MEDLINE=98445411; PubMed=9770526;
RA Martham R.B., Wang W.C., Weinstein A.E., Wang Z., Munoz A.,
Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H.,
Yu X.F.;
RT "Patterns of HIV-1 evolution in individuals with differing rates of
RT CD4 T cell decline.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).
DR EMBL; AF089191; AAC78901.1; -.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 94 AA; 10797 MW; 22C21B4D098FA672 CRC64;

Query Match 27.6%; Score 53; DB 15; Length 94;
Best Local Similarity 43.5%; Pred. No. 46;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 4 HPYGTLPVGPPEFKTLRVONL 26
DB 34 HTVRKIPIGPSSFTYTGVRGDI 56

RESULT 9
09PU31 PRELIMINARY; PRT; 714 AA.
ID 09PU31
AC 09PU31;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE C-ski protein.
GN C-SKI.
OS Oreochromis aureus (Israeli tilapia).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Perciformes; Labroidae;
OC Cichlidae; Oreochromis.
OX NCBI_TaxId=47969;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang C.J., Lin J.Y., Tsai H.J.;
RT "Two distinct c-ski proto-oncogene cDNAs of fish, tilapia (Oreochromis
RT aurea).";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ012012; CAB58126.1; -.
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
SQ SEQUENCE 714 AA; 81076 MW; F2F8AB6607CCC72F CRC64;

Query Match 27.6%; Score 53; DB 13; Length 714;
Best Local Similarity 35.3%; Pred. No. 46;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVONLGCYVSG 33
DB 63 FVMPGPELFVPSDRSTERCETVLEGETISCFVVG 96

RESULT 10

DB 39 IPGGRAFYTTGRIGNI 56

RESULT 12

ID 09YT23 PRELIMINARY; PRT; 94 AA.

AC 09YT23;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE Envelope glycoprotein (Fragment).

GN ENV.

OS Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OC NCBI_TaxID=11676;

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S4V4-7;

RX MEDLINE=98445411; PubMed=9770526;

RA Marham R.B., Wang W.C., Weistein A.E., Wang Z., Munoz A., Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H., Yu X.F.;

RT "Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline."

RT Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).

RL EMBL; AF089188; AAC78898.1;

DR EMBL; AF089188; AAC78898.1;

DR InterPro: IPR000777; GP120.

DR Pfam: PF00516; GP120.1.

DR AIDS; Coat protein; Glycoprotein.

KW NON_TER

FT NON_TER

SQ SEQUENCE 94 AA; 10781 MW; 249420C1B250B370 CRC64;

Query Match 27.1%; Score 52; DB 13; Length 94;

Best Local Similarity 50.0%; Pred. No. 6.4;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 67 PVGPGPFLPSDRSTERCIVLEGRTISCFVGG 100

10 PVGPGPFLPSDRSTERCIVLEGRTISCFVGG 33

RESULT 11

ID 09YT25 PRELIMINARY; PRT; 94 AA.

AC 09YT25;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Envelope glycoprotein (Fragment).

GN ENV.

OS Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OC NCBI_TaxID=11676;

OX NCBI_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S4V4-4;

RX MEDLINE=98445411; PubMed=9770526;

RA Marham R.B., Wang W.C., Weistein A.E., Wang Z., Munoz A., Templeton A., Margolick J., Vlahov D., Quinn T., Farzadegan H., Yu X.F.;

RT "Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline."

RT Proc. Natl. Acad. Sci. U.S.A. 95:12568-12573(1998).

RL EMBL; AF089185; AAC78895.1;

DR EMBL; AF089185; AAC78895.1;

DR InterPro: IPR000777; GP120.

DR Pfam: PF00516; GP120.1.

DR AIDS; Coat protein; Glycoprotein.

FT NON_TER

SQ SEQUENCE 94 AA; 10750 MW; DF7420C1A6179C68 CRC64;

Query Match 27.1%; Score 52; DB 15; Length 94;

Best Local Similarity 50.0%; Pred. No. 6.4;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

DB 9 LPVGPGEFFKTLRVONL 26

9 LPVGPGEFFKTLRVONL 26

DB 39 IPGGRAFYTTGRIGNI 56

RESULT 13

ID 09UWCI PRELIMINARY; PRT; 359 AA.

AC 09UWCI;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE CFU99.

GN Gallus gallus (Chicken).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Kaneko M., Nishihara S., Kitano T., Narimatsu H., Saitou N.;

RT "The evolutionary history of glycosyltransferase genes."

RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

RL EMBL; AB035906; BAB82489.1;

DR EMBL; AB035906; BAB82489.1;

DR InterPro: IPR001503; GT_10.

DR Pfam: PF00852; Glyco_transf_10;

SQ SEQUENCE 359 AA; 42077 MW; 2A19DACD1F49B6C8 CRC64;

Query Match 27.1%; Score 52; DB 15; Length 94;

Best Local Similarity 42.9%; Pred. No. 30;

Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

DB 9 LPVGPGEFFKTLRVONL 26

39 IPGGRAFYTTGRIGNI 56

RESULT 13

ID 09UWCI PRELIMINARY; PRT; 359 AA.

AC 09UWCI;

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE CFU99.

GN Gallus gallus (Chicken).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Kaneko M., Nishihara S., Kitano T., Narimatsu H., Saitou N.;

RT "The evolutionary history of glycosyltransferase genes."

RT Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

RL EMBL; AB035906; BAB82489.1;

DR EMBL; AB035906; BAB82489.1;

DR InterPro: IPR001503; GT_10.

DR Pfam: PF00852; Glyco_transf_10;

SQ SEQUENCE 359 AA; 42077 MW; 2A19DACD1F49B6C8 CRC64;

Query Match 27.1%; Score 52; DB 13; Length 359;

Best Local Similarity 42.9%; Pred. No. 30;

Matches 12; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

DB 5 PYGTLPVGPGEFFKTLRVONLGCYVVS 32

5 PYGTLPVGPGEFFKTLRVONLGCYVVS 32

Db 167 PYGSMIVGTGATFYENPSKENILVCWVS 194

RESULT 14

O979H6

AC O979H6; PRELIMINARY; PRT; 461 AA.

DT 01-OCT-2001 (TREMBlrel. 18, Created)

DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Shikimate transporter.

GN TV1185 OR TVG1213986.

OS Thermoplasma volcanium.

OC Archaeae; Euryarchaeota; Thermoplasmatia; Thermoplasmales;

OC Thermoplasmataceae; Thermoplasma.

OX NCB1_TaxID=50339;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GSSI / DSM 4299 / JCM 9571;

RX MEDLINE=20570466; PubMed=11121031;

RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,

RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,

RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.,

RT "Archaeal adaptation to higher temperatures revealed by genomic

RT Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).

DR EMBL; AP000995; BAB60327.1; -.

DR InterPro; IPR003662; sub.transporter.

DR Pfam; PF00083; sugar tr. 1.

DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 461 AA; 50273 MW; F4BA80954FD218F4 CRC64;

Query Match 27.1%; Score 52; DB 17; Length 461;

Best Local Similarity 36.7%; Pred. No. 39;

Matches 11; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

OY 5 PYGTLPGVGPGEFKTTLRVQNLGCVVSG 34

DB 371 PWGIVPVYLSERFKAIVRASGVGFGVSSGI 400

RESULT 15

O8VIL5

AC O8VIL5; PRELIMINARY; PRT; 725 AA.

DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Ski proto-oncogene.

GN Ski.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCB1_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=129/SV.

RA Chen Y., Berk M., Chen H., Stavnezer E., Colmenares C.;

RT "Mouse Ski proto-oncogene cDNA."

RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF435852; AAL30825.1; -.

DR InterPro; IPR003380; Transform_Ski.

DR Pfam; PF02437; Ski_Smc.1.

SQ SEQUENCE 725 AA; 80119 MW; 1BFD05C38519505C CRC64;

Query Match 27.1%; Score 52; DB 11; Length 725;

Best Local Similarity 35.3%; Pred. No. 66;

Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

OY 10 PYGPGPERF-----KTLRVQNLGCVVSG 33

DB 82 PVLPGPFPMPSDRSTERCEIVLEGISCFVVG 115

Search completed: July 16, 2003, 07:55:28
Job time : 6.59353 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 3.88889 Seconds

(without alignments)
1199.256 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHAPYGTLPVGPGEFXTTLRVONLCGYVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	100.0	1152	22 AAB64657	Human secreted pro
2	192	100.0	1152	22 AAB64658	Human secreted pro
3	192	100.0	1167	21 AAY32242	Human integrin sub
4	192	100.0	1167	22 AAB64584	Human integrin sub
5	142	74.0	1132	21 AAY32243	Human integrin sub
6	69	35.9	1188	22 AAU14231	Human novel protei
7	69	35.9	1188	22 AAU14467	Human novel protei
8	69	35.9	1188	22 AAB50085	Human A259. Homo
9	69	35.9	1188	23 AAU10551	Human A259 polypep
10	59	30.7	1034	21 AAB25590	Protein encoded by

11	59	30.7	1188	22 AAB30929	Amino acid sequenc
12	59	30.7	1189	21 AAB25582	ITGAL1 protein enc
13	59	30.7	1189	22 AAB212949	Novel human diagno
14	56	29.2	347	12 AAR11069	12D3 antigen sequ
15	56	29.2	545	23 AAB72288	Murine protein iso
16	56	29.2	688	23 AAB72300	Rat protein isolat
17	56	29.2	696	23 AAB72289	Rat protein isolat
18	56	29.2	1188	22 AAB50087	Murine A259. Mus
19	56	29.2	1188	22 AAB50087	Murine A259 polype
20	55	28.6	210	22 AAU62597	Propionibacterium
21	54	28.1	257	22 AAB20619	Novel human diagno
22	54	28.1	391	22 AAB20618	Novel human diagno
23	53.5	27.9	349	22 AAB21433	Novel human diagno
24	52.5	27.3	267	22 AAB28949	Novel human diagno
25	52	27.1	346	12 AAR11068	12D3 antigen sequ
26	52	27.1	728	22 AAB01251	Human polypeptide
27	52	27.1	728	22 AAB01251	Human c-ski oncopr
28	52	27.1	750	12 AAR14048	Human c-ski oncopr
29	52	27.1	750	12 AAR15160	FB29 chicken c-ski
30	52	27.1	750	12 AAR10458	FB29 chicken c-ski
31	52	27.1	750	22 AAB01022	Chicken c-ski FB29
32	52	27.1	750	22 AAB01025	Alternative versio
33	52	27.1	752	22 AAB42037	Human polypeptide
34	52	27.1	1111	23 AAB017108	Human c-ski oncopr
35	51	26.6	106	23 AAB05801	Murine multi-copper
36	51	26.6	741	23 AAU74438	Human multi-copper
37	51	26.6	846	15 AAR47474	Mouse protein sequ
38	51	26.6	890	15 AAR60252	Potato sucrose pho
39	51	26.6	890	22 AAB78052	Human T lymphocyte
40	50.5	26.3	462	22 AAU30882	Novel human secret
41	50	26.0	79	23 AAB42461	Human ovarian anti
42	50	26.0	496	23 AAU74443	Human protein sequ
43	50	26.0	613	21 AAB19339	Amino acid sequenc
44	50	26.0	674	23 AAU74445	Human protein sequ
45	50	26.0	714	23 AAU74441	Human protein sequ

ALIGNMENTS

RESULT 1	AAAB64657	standard; Protein; 1152 AA.
ID	AAAB64657	standard; Protein; 1152 AA.
XX	AAAB64657	
AC	AAAB64657	
XX	AAAB64657	
DT	22-MAR-2001	(first entry)
XX	22-MAR-2001	
DE	Human secreted protein BLAST search protein SEQ ID NO: 167.	
XX	Human secreted protein BLAST search protein SEQ ID NO: 167.	
XX	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;	
KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KW	neurological disease; infection; human; secreted protein.	
OS	Homo sapiens.	
XX	Homo sapiens.	
XX	WO200077197-A1.	
XX	21-DEC-2000.	
XX	21-DEC-2000.	
PF	01-JUN-2000; 2000WO-US14934.	
XX	01-JUN-2000; 2000WO-US14934.	
XX	11-JUN-1999; 99US-0138599.	
PR	11-JUN-1999; 99US-0138599.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	(ROSE/) ROSEN C A.	
PI	Rosen CA, Ruben SM, Komatsoulis GA;	
XX	Rosen CA, Ruben SM, Komatsoulis GA;	
DR	WPI; 2001-032312/04.	
XX	WPI; 2001-032312/04.	

PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; Page 543-546; 558pp; English.

CC The invention relates to the isolation of genes AAF32757-F32803 encoding
 CC the human secreted proteins AAB64549-B64594. The sequence is a search
 CC result from a BLASTX homology search. The genes and proteins are useful
 CC for preventing, ameliorating or treating medical conditions, e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer, and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemia; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.

SQ Sequence 1152 AA;
 Query Match 100.0%; Score 192; DB 22; Length 1152;
 Best Local Similarity 100.0%; Pred. No. 6.2e-18;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 35
 |||||
 DB 952 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 986

RESULT 2
 AAB64658
 ID AAB64658 standard; Protein; 1152 AA.

AC AAB64658;
 XX
 DT 22-MAR-2001 (first entry)
 XX
 DE Human secreted protein BLAST search protein SEQ ID NO: 168.
 XX
 KW Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200077197-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-US14934.
 XX
 PR 11-JUN-1999; 99US-0138599.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 DR WPI; 2001-032312/04.
 XX
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; Page 547-551; 558pp; English.
 XX
 CC The invention relates to the isolation of genes AAF32757-F32803 encoding

CC the human secreted proteins AAB64549-B64594. The sequence is used as a
 CC query sequence for doing BLASTX searches to identify homologous
 CC sequences. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal
 CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
 CC disease, allergies, autoimmune haemolytic anaemia, autoimmune
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
 CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
 CC disorders such as myocardial ischaemia; (d) wound healing; (e)
 CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections.

SQ Sequence 1152 AA;
 Query Match 100.0%; Score 192; DB 22; Length 1152;
 Best Local Similarity 100.0%; Pred. No. 6.2e-18;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 35
 |||||
 DB 952 YEHPHYGTLPGPGPEFKTTLRVONLGCYVSGLI 986

RESULT 3
 AAY32242
 ID AAY32242 standard; Protein; 1167 AA.

AC AAY32242;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Human integrin subunit alpha-10.
 XX
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KW fibroblast; vaccine; marker.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /note= "signal peptide"
 FT 23..1145
 FT /note= "mature protein"
 FT 23..1120
 FT /note= "extracellular domain"
 FT 1121..1145
 FT /note= "transmembrane domain"
 FT 1122..1167
 FT /note= "cytoplasmic domain, specifically claimed
 FT in Claim 21"
 FT Domain 162..359
 FT /note= "I-domain"
 FT 494..502
 FT /note= "cation binding site motif"
 FT 558..566
 FT /note= "cation binding site motif"
 FT 620..628
 FT /note= "cation binding site motif"
 FT 98
 FT /note= "N-glycosylated"
 FT 336
 FT /note= "N-glycosylated"
 FT 364
 FT /note= "N-glycosylated"
 FT Modified-site 733

FT	/note= "N-glycosylated"
FT	Modified-site 839
FT	/note= "N-glycosylated"
FT	Modified-site 921
FT	/note= "N-glycosylated"
FT	Modified-site 1018
FT	/note= "N-glycosylated"
FT	Modified-site 1039
FT	/note= "N-glycosylated"
XX	
PN	WO951639-A1.
PD	
PF	14-OCT-1999.
PF	
PF	31-MAR-1999; 99WO-SE00544.
PR	
PR	02-APR-1998; 98SE-.0001164.
PR	28-JAN-1999; 99SE-.0000319.
PA	(ACTI-) ACTIVE BIOTECH AB.
XX	
P1	Lundgren-Akerlund E;
DR	WPI; 2000-052639/04.
DR	N-PSDB; AAZ34719.
PT	
PT	New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation -
XX	
PS	Claim 1; Fig 6; 90DP; English.
XX	
CC	This sequence represents novel human chondrocyte integrin subunit alpha-10 (ISa10). A splice variant or isolated in AMY32243. The invention relates to a recombinant or isolated integrin heterodimer comprising the alpha10 subunit in association with subunit beta (especially beta-1). The heterodimer and the subunit alpha-10 can be used as markers or targets of all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also be used: for treating pathological conditions involving ISa10, such as damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for detecting the formation of cartilage during embryonal development, physiological or therapeutic repairation of cartilage, or detecting regeneration of cartilage or chondrocytes during transplantation of cartilage or chondrocytes; for selection and analysis or for sorting, isolating or purification of chondrocytes and for in vitro studies of differentiation of chondrocytes and as a target for anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or other tissues where adhesion impairs the function of the tissue (all claimed). ISa10 binding entities can be used to determine the differentiation-state of cells during embryonic development, angiogenesis or development of cancer, in pathological conditions such as rheumatoid arthritis, osteoarthritis or cancer, in tissue regeneration (claimed) . A vaccine comprising the integrin of cartilage (claimed) . A vaccine comprising the integrin heterodimer or subunit alpha-10 is also claimed. ISa10 polynucleotides, vectors, host cells and methods of producing recombinant ISa10 are also claimed.
Sequence	1167 AA;
150	

	Query Match	100.0%	Score 192;	DB 21;	Length 1167;
	Best Local Similarity	100.0%	Pred. No. 6,3-18;		
Matches	35; Conservative	0;	Mismatches	0;	Indels
Gaps					0;
Oy	1 YEYHPYGTLPVGPGEPEKRTLRPVQNLGCYVNSGLI	35			
Dd	YEYHPYGTLPVGPGEPEKRTLRPVQNLGCYVNSGLI	986			
RESULT 4					
AAB64584					

ID	AA064584	standard; Protein: 1167 AA.
AC	AA064584;	
XX		
DT	22-MAR-2001	(first entry)
XX		
DE	Human secreted protein #37.	
XX		
KM	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KM	antiallergic; hepatotrophic; antidiabetic; antiinflammatory; antiviral;	
KM	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KM	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;	
XX	neurological disease; infection; human; secreted protein.	
XX		
OS	Homo sapiens.	
PN	WO200077197-A1.	
XX		
PD	21-DEC-2000.	
XX		
PF	01-JUN-2000; 2000WO-US14934.	
XX		
FR	11-JUN-1999; 99US-0138599.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
PA	(ROSE/) ROSEN C A.	
XX		
PI	Rosen CA, Ruben SM, Komatsoulis GA;	
XX		
DR	WPI; 2001-032312/04.	
DR	N-PSDB; AAF32793.	
XX		
PT	Isolated nucleic acid molecule encoding a human secreted protein is	
PT	used in preventing, treating or ameliorating a medical condition -	
XX		
PS	Claim 11; Page 496-500; 558pp; English.	
XX		
CC	Sequences AA064549-864594 represent the amino acid sequences of 47	
CC	human secreted proteins encoded by the genes AAF32757-F32803. The genes	
CC	and proteins are useful for preventing, ameliorating or treating medical	
CC	conditions, e.g. by protein or gene therapy. The genes are isolated from	
CC	a range of human tissues disclosed in the specification. The nucleic	
CC	acids, proteins, antibodies and (ant)agonists are useful in the	
CC	diagnosis, treatment and prevention of: (a) cancer, e.g. breast and	
CC	ovarian cancer, and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,	
CC	Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
XX		
Q0	Sequence 1167 AA;	

[illegible]

DE Human integrin subunit alpha-10 splice variant.

XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KW fibroblast; vaccine; marker; splice variant.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..22 /note="signal peptide"
 FT Protein 23..1132
 FT /note="mature protein"

XX MO951639-A1.

XX 14-OCT-1999.

XX 31-MAR-1999; 99WO-SE00544.

XX 02-APR-1998; 98SE-0001164.

XX 28-JUN-1999; 99SE-0000319.

XX (ACT1-) ACTIVE BIOTECH AB.

XX Lundgren-Akerlund E;

XX WPI; 2000-052639/04.

XX N-PSDB; AAS24720.

PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation

XX Claim 1; Page 43-48; 90pp; English.

XX This sequence represents a splice variant of novel human
 CC chondrocyte integrin subunit alpha-10 (ISa10). It is identical to
 CC ISa10 (see AAI32242) except for deletion of amino acids 975-986. The
 CC invention relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The heterodimer, subunit alpha-10 or splice
 CC variant can be used as a marker or target of all types of cells, e.g.
 CC of chondrocytes, osteoblasts and fibroblasts. They can also be used:
 CC for treating pathological conditions involving ISa10, such as
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage,
 CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). ISa10 binding entities can be used to
 CC determine the differentiation-state of cells during embryonic
 CC development, angiogenesis or development of cancer, in pathological
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
 CC in tissue regeneration or in therapeutic and physiological repair
 CC of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. ISa10
 CC polynucleotides, vectors, host cells and methods of producing
 CC recombinant ISa10 are also claimed.

XX Sequence 1132 AA;

Query Match 74.0%; Score 142; DB 21; Length 1132;

Best Local Similarity 83.3%; Pred. No. 5 6e-11;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 YEVHPYGLPVGPGPEFKTLRTNNASCIV 30

DB 952 YEVHPYGLPVGPGPEFKTLRTNNASCIV 981

RESULT 6
 ID AAU14231 standard; Protein; 1188 AA.

XX AAU14231;

XX 24-OCT-2001 (first entry)

XX Human novel protein #102.

XX Human; novel protein; Antianemic; osteopathic; antiinflammatory;
 KW immunomodulatory; cytoarrestic; neuroprotective; vulnery; nootropic;
 KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
 KW antibacterial; antiallergic; dermatological; haemostatic; antiaesthetic;
 KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
 KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
 KW tissue regeneration; immune disorder.

XX Homo sapiens.

XX WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Demanac RT;

XX WPI; 2001-451939/48.

XX N-PSDB; AAS22536.

PT Isolated polypeptides useful for treating anti-inflammatory diseases,
 PT nervous system disorders, and for regenerating bone and cartilage -
 XX Example 4; Page 578-581; 894pp; English.

XX The invention relates to polynucleotides encoding novel human
 CC proteins or their active domains. The polypeptides, polynucleotides and
 CC antibodies raised against the polypeptides are used in a method of
 CC treatment of a mammal and prevention of disorders caused by the aberrant
 CC protein expression or activity. The polypeptides can be used as
 CC molecular weight markers, food supplements, and in antibody production.
 CC The polypeptides are used to identify compounds which bind to the
 CC polypeptides. Polynucleotides of the invention are used as probes and
 CC primers, for sequencing, for chromosome or gene mapping, in the
 CC production of recombinant proteins, and in generating anti-sense DNA or
 CC RNA and in gene therapy. Polypeptides of the invention can be used to
 CC target drugs to a tumour, in assays to determine biological activity, to
 CC raise antibodies/elicit an immune response, to determine quantitative
 CC protein levels, as tissue markers, and to isolate receptors or ligands.
 CC Polypeptides of the invention may also be useful in creating platelet
 CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
 CC ligament and/or nerve tissue, wound healing, treating burns, promoting
 CC the proliferation, differentiation and survival of stem cells, as a
 CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
 CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
 CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
 CC fungal infection or from autoimmunity, cancer, allergy, asthma,
 CC graft-versus-host disease, eczema, haemophilia, thrombosis,
 CC anti-inflammatory diseases, nervous system disorders, and infection.
 CC The present sequence represents a protein of the invention.

XX Sequence 1188 AA;

Query Match 35.9%; Score 69; DB 22; Length 1188;

PT diagnosis of fibrosis, e.g. of the liver
 PS Claim 8; Fig 1; 164pp; English.
 XX
 CC The present sequence is human integrin alpha subunit, A259. A259 is
 CC homologous with the alpha1 and alpha10 integrin subunits and is
 CC overexpressed in fibrosis. A259 is implicated in regulation of
 CC proliferation, differentiation and/or function of many different cell
 CC types. Inhibitors of A259 activity are useful for the treatment of liver
 CC disease, particularly fibrosis, and also fibrosis in other organs
 CC (specifically lung and kidney). In addition, A259 can be used for
 CC treatment and prevention of cancer, osteoporosis, acute myeloid
 CC leukaemia, HIV infection, and rheumatoid arthritis.
 CC
 SO Sequence 1188 AA:
 Query Match 35.9%; Score 69; DB 22; Length 1188;
 Best Local Similarity 40.5%; Pred. No. 0.84;
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;
 QY 1 YEWHPYGTLP--VGPGPEFKTLRVNIGCYVSGLI 35
 968 YEVENSSLERYDGIQPFSCIFRIQNLGLPFIHGM 1004
 DB
 RESULT 9
 AAU10551 standard; Protein; 1188 AA.
 XX
 AC AAU10551;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Human A259 polypeptide.
 XX
 KW Human; A259; integrin alpha subunit; integrin alpha 10; secreted protein;
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;
 KW immune related disease; apoptotic disorder; neuronal tissue disease;
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
 KW antidiabetic; antinaeumatic; antiallergic; antiaesthetic; dermatological;
 KW antidiabetic; anticonvulsant; antiparkinsonian.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..22
 FT /note= "Signal peptide"
 FT 1..1141
 FT /note= "Extracellular domain"
 FT 23..1188
 FT /note= "Mature human A259"
 FT 37..90
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 115..157
 FT /note= "Integrin alpha repeat domain"
 FT 164..345
 FT /note= "domain or Von Willebrand Factor type A domain"
 FT 367..392
 FT /note= "Integrin alpha repeat domain"
 FT 421..472
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 476..532
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 538..593
 FT /note= "Integrin alpha repeat domain. The specification
 FT states that this domain exists in human A549"
 FT 600..654
 FT /note= "Integrin alpha repeat domain. The specification

FT
 FT Domain 1142..1164 states that this domain exists in human A549"
 FT /note= "Transmembrane domain"
 FT 1165..1188
 FT Domain /note= "Cytoplasmic domain"
 FT
 PN WO200181414-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 27-APR-2001; 2001WO-US13516.
 XX
 PR 27-APR-2000; 2000US-0561263.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Pan Y, Lora J;
 XX
 DR WPI; 2002-041397/05.
 XX
 DR N-PSDB; AAS16873.
 XX
 PT New A259 nucleic acids and polypeptides, which comprise integrin alpha
 PT subunit, useful for diagnosing, preventing or treating e.g. liver
 PT disease, kidney or lung fibrosis, cancers, blood disorders or immune
 PT related diseases
 XX
 PS Claim 9; Fig 1; 168pp; English.
 XX
 CC The invention relates to human and murine A259 nucleic acid molecules
 CC which encode secreted proteins with homology to integrin alpha subunits,
 CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
 CC are useful for treating liver disease or fibrosis, particularly kidney
 CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
 CC useful for diagnosing, preventing or treating cartilage and bone
 CC associated disorders (such as bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
 CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as
 CC acute myeloid leukaemia, haemophilia, anaemia and thalassaemia), immune
 CC related diseases (such as HIV, viral infections, cancers, T cell
 CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
 CC asthma and psoriasis), apoptotic disorders (such as systemic lupus
 CC erythematosus and insulin-dependent diabetes mellitus), diseases of the
 CC neuronal tissues (such as epilepsy and muscular dystrophy) and
 CC neurodegenerative diseases (such as Parkinson's disease and Huntington's
 CC disease). This sequence represents the human A259 polypeptide.
 CC
 SQ Sequence 1188 AA;
 Query Match 35.9%; Score 69; DB 23; Length 1188;
 Best Local Similarity 40.5%; Pred. No. 0.84;
 Matches 15; Conservative 6; Mismatches 14; Indels 2; Gaps 1;
 QY 1 YEWHPYGTLP--VGPGPEFKTLRVNIGCYVSGLI 35
 968 YEVENSSLERYDGIQPFSCIFRIQNLGLPFIHGM 1004
 DB
 RESULT 10
 AAB25590
 ID AAB25590 standard; Protein; 1034 AA.
 XX
 AC AAB25590;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Protein encoded by human secreted protein gene #7 clone HOHB569.
 XX
 KW Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
 KW antirheumatic; dermatological; antiproliferative; antiatherosclerotic;
 KW anticancer; vulnary; antiviral; antibacterial; antifungal;
 KW immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;

KM cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
 XX
 OS Homo sapiens.
 XX
 PN WO200029435-A1.
 PD 25-MAY-2000.
 XX
 PF 27-OCT-1999; 99WO-US25031.
 XX
 PR 28-OCT-1998; 98US-0105971.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ruben SM, Olsen HS, Young PE, Kenny JT, Moore PA, Wei Y,
 PI Greene JM;
 XX
 DR WPI: 2000-387742/33.
 XX
 PF Isolated nucleic acid molecules encoding human secreted proteins are
 PF used for the prevention, amelioration and treatment of autoimmune,
 PF inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PF wounds, and infectious diseases -
 XX
 PS Claim 1; Page 678-682; 803pp; English.
 XX
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAB0606-A80623 encode the 12 secreted protein sequences given in
 CC AAB5576-B5593. The human secreted proteins have various activities
 CC dependant on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antiarthritic; antirheumatic; dermatological;
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraproteinemia and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #7 and protein
 CC sequences are represented in sequences AAB0612 and AAB25582. Secreted
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
 CC AAB0652-A80661 represent genes which are related to the secreted protein
 CC gene#7.
 CC
 XX
 SQ Sequence 1034 AA;
 XX
 QY Query Match 30.7%; Score 59; DB 21; Length 1034;
 XX Best Local Similarity 41.7%; Pred. No. 18;
 DB Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 XX
 12 GGPPEFKTTLRVNLCGVVSGLI 35
 DB 981 GIGPFFSCIFRIQNLGFLPIHGM 1004
 XX
 RESULT 11
 AAB30929
 ID AAB30929 standard; Protein; 1188 AA.
 XX
 AC AAB30929;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Amino acid sequence of a human alpha1 integrin chain.
 XX
 KW Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;

KM osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
 KW wound healing; trauma; rheumatoid arthritis; osteoarthritis;
 XX osteoporosis; cartilage damage; bone damage; cartilage.
 XX
 OS Homo sapiens.
 XX
 PN WO200075187-A1.
 PD 14-DEC-2000.
 XX
 PF 31-MAY-2000; 2000WO-SE01135.
 XX
 PR 03-JUN-1999; 99SE-0002056.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 PI Gullberg D;
 PI
 XX
 DR WPI: 2001-071061/08.
 DR N-PsDB; AAC68671.
 XX
 PF Integrin subunit alpha 11 or integrin heterodimer comprising subunit
 PF alpha 11 in association with subunit beta, useful for treating muscle
 PF dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis -
 XX
 PS Disclosure: Fig 2a-c; 79pp; English.
 XX
 CC The present sequence represents a human integrin subunit, designated
 CC alpha11. The alpha11 polynucleotide and polypeptide are useful as
 CC markers of cell target molecules, such as fibroblasts, muscle cells,
 CC chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
 CC They are also used for determining the differential-stage of cells
 CC during differentiation, development in pathological conditions, in
 CC tissue regeneration, in transplantation or in therapeutic and
 CC physiological repair of tissues. The pathological conditions involving
 CC subunit alpha11 are selected from damage of cells, muscle dystrophy,
 CC fibrosis, wound healing, trauma, rheumatoid arthritis, osteoarthritis
 CC and osteoporosis, damage of cartilage and bone, and cartilage and bone
 CC diseases. The polypeptide is useful for detecting the formation of
 CC cartilage during embryonic development, for detecting physiological
 CC therapeutic repair of cartilage and muscle, for selection and analysis,
 CC or for sorting, isolating or purification of chondrocytes and muscle
 CC cells, for detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes during transplantation of
 CC cartilage or chondrocytes, respectively, or of muscle or muscle cells
 CC during transplantation of muscle or muscle cells, respectively, and for
 CC studies of differentiation or chondrocytes or muscle cells.
 CC
 XX
 SQ Sequence 1188 AA;
 XX
 QY Query Match 30.7%; Score 59; DB 22; Length 1188;
 XX Best Local Similarity 41.7%; Pred. No. 21;
 DB Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
 XX
 12 GGPPEFKTTLRVNLCGVVSGLI 35
 DB 981 GIGPFFSCIFRIQNLGFLPIHGM 1004
 XX
 RESULT 12
 AAB25582
 ID AAB25582 standard; Protein; 1189 AA.
 XX
 AC AAB25582;
 XX

DT 21-NOV-2000 (first entry)

XX ITGA11 protein encoded by human secreted protein gene #7.

XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
XX antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
XX anticancer; vulnery; antiviral; antibacterial; antifungal;
XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
XX Crohn's disease; nephritis; hyperproliferative disorder;
XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
XX melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.

OS Homo sapiens.

PN WO200029435-A1.

PD 25-MAY-2000.

PF 27-OCT-1999; 99WO-US25031.

PR 28-OCT-1998; 98US-0105971.

PA (HUMA-) HUMAN GENOME SCI INC.

PI N1 J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;

PI Greene JM;

DR WPI, 2000-387742/33.

DR N-PSDB; AAA80612.

XX Isolated nucleic acid molecules encoding human secreted proteins are
XX used for the prevention, amelioration and treatment of autoimmune,
XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases -

XX Claim 1; Figure 19A-F; 803pp; English.

XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given
XX in AAA80606-AB80623 encode the 12 secreted protein sequences given in
XX AAB25576-B25593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant;
XX anti-inflammatory; antirheumatic; antirheumatic; dermatological;
XX antiproliferative; antiproliferative; anticancer; vulnery;
XX antiviral; antibacterial; and antifungal activity. The proteins,
XX polypeptides, agonists and antagonists may be used to treat prevent
XX and/or diagnose various disease, disorders and conditions examples of
XX which include: immune disorders e.g. Addison's disease, rheumatoid
XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
XX hyperproliferative disorders such as paraproteinemia and purpura;
XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
XX sequences may also be used in wound healing and the treatment of
XX infectious diseases. The human secreted protein gene #7 and protein
XX sequences are represented in sequences AAA80612 and AAB25582. Secreted
XX protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
XX AAA80605-AB80661 represent genes which are related to the secreted protein
XX gene#7.

XX Sequence 1189 AA;

Query Match 30.7%; Score 59; DB 21; Length 1189;

Best Local Similarity 41.7%; Pred. No. 21;

Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTLRVQNLGCVVSGLI 35

DB 981 GIGPPSCIFRIQNLGLFPIHGM 1004

RESULT 13

ABG12949

XX ABG12949 standard; Protein; 1189 AA.

AC ABG12949;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #12940.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI, 2001-639362/73.

DR N-PSDB; AAS77136.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -

XX Claim 20; SEQ ID NO 43308; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1189 AA;

Query Match 30.7%; Score 59; DB 22; Length 1189;

Best Local Similarity 41.7%; Pred. No. 21;

Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 12 GPGPEFKTLRVQNLGCVVSGLI 35

DB 981 GIGPPSCIFRIQNLGLFPIHGM 1004

RESULT 14

AAR11069

```

ID  AAR1069 standard; Protein; 347 AA.
XX  AAR1069;
XX  23-MAY-1991 (first entry)
DT
XX  12D3 antigen sequence deduced from cDNA of pB101.
DE
XX  Babesia; red water fever; cattle; antibodies.
XX  Babesia bovis.
XX  Babesia bovis.
XX  Key Location/Qualifiers
XX  Peptide 1..21
XX  Protein /label= signal sequence
XX  22..346
XX  /label= mature protein
XX  EP417524-A.
XX  20-MAR-1991.
XX  23-AUG-1990; 90EP-0116165.
XX  23-AUG-1989; 89AU-0005902.
XX  (CSIR ) COMMONWEALTH SCIENT ORG.
XX  Riddles PW, Aylward JH, Wright IG;
XX  WPI; 1991-081822/12.
DR  N-PSDB; AAQ10952.
XX
XX  New antigen from Babesia for protective vaccine - used with
XX  derived antibodies, DNA sequences and oligo:nucleotide probes,
XX  for immuno-diagnosis.
XX
XX  Claim 13; Fig 9; 40pp; English.
XX
XX  The sequence was deduced from a clone isolated from a cDNA library
XX  prep'd from bovine erythrocytes infected with Babesia bigemina.
XX  The DNA can be used to construct a vector for expression of the
XX  antigen which can be used in vaccines to protect cattle from red
XX  water fever.
XX  See also AAR1068.
XX
XX  Sequence 347 AA;
SQ
Query Match 29.2%; Score 56; DB 12; Length 347;
Best Local Similarity 44.7%; Pred. No. 15;
Matches 17; Conservative 2; Mismatches 13; Indels 6; Gaps 3;

OY 1 YEYHRY-GTLPVGP---GPEFKTLR--VQNLGCVWS 32
   ||| | | | | | | | | | | | | | | |
DB 67 YIVNRYIGRCPIGTTCGPEFDATLGTGADPSGCVTVS 104

RESULT 15
ABR72288
ID  ABR72288 standard; Protein; 545 AA.
XX  ABR72288;
XX  04-APR-2002 (first entry)
DT
XX  Murine protein isolated from skin cells SEQ ID NO: 500.
XX
XX  Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
XX  developmental defect; inflammatory disease; dermatological; vulnerary;
XX  immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX  Mus sp.
XX

```

```

EN  WO200190357-A1.
XX
XX  29-NOV-2001.
XX
XX  24-MAY-2001; 2001WO-N200099.
XX
XX  24-MAY-2000; 2000US-206650P.
XX  25-JUL-2000; 2000US-221232P.
XX  (GENE-) GENESIS RES & DEV CORP LTD.
XX
XX  Watson JD, Strachan L, Sleeman M, Onrust R, Morrison JG, Kumble KD;
XX  WPI; 2002-122020/16.
XX
XX  New polynucleotides and polypeptides encoded by the polynucleotides
XX  isolated from skin cells, useful for treating skin wounds, cancers,
XX  growth and developmental defects, inflammatory diseases, or for
XX  modulating immune responses
XX
XX  Claim 4; Page 305-306; 466pp; English.
XX
XX  The present invention provides the protein and coding sequences of cDNAs
XX  isolated from human, murine and rat skin cell libraries. The sequences
XX  can be used in the development of therapeutic agents useful in the
XX  treatment of skin diseases, including skin wounds, cancer, growth
XX  defects, developmental defects and inflammatory diseases. The proteins
XX  have important roles in the induction of hair growth, cell proliferation
XX  and cell-cell interaction, in maintaining tissue integrity, in wound
XX  healing and in modulating immune responses. The present sequence is a
XX  polypeptide of the invention.
XX
XX  Sequence 545 AA;
SQ
Query Match 29.2%; Score 56; DB 23; Length 545;
Best Local Similarity 41.7%; Pred. No. 24;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 12 GPGEFKTLRVQNLGCVWSGLI 35
   ||| | | | | | | | | | | | | | | |
DB 338 GIGPPFNCVFNQNLGFPPIHGVW 361

Search completed: July 16, 2003, 07:52:07
Job time : 5.88889 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:55:38 ; Search time 2.5316 Seconds
(without alignments)
1639.577 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHPHYGTLPGPGEFRTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.dep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.dep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.dep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.dep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.dep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.dep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.dep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.dep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.dep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.dep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.dep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.dep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.dep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.dep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	30.7	1034	9 US-09-984-130-43	Sequence 43, Appl
2	59	30.7	1034	9 US-09-836-353A-43	Sequence 43, Appl
3	59	30.7	1189	9 US-09-984-130-35	Sequence 35, Appl
4	59	30.7	1189	9 US-09-836-353A-35	Sequence 35, Appl
5	57.5	28.9	3338	9 US-10-156-761-8464	Sequence 8464, Ap
6	56	28.2	545	9 US-09-866-050A-500	Sequence 500, App
7	56	28.2	688	9 US-09-866-050A-624	Sequence 624, App
8	52	27.1	696	9 US-09-866-050A-501	Sequence 501, App
9	52	27.1	308	9 US-10-156-761-13191	Sequence 13191, A
10	52	27.1	728	10 US-09-753-831-2	Sequence 2, Appli
11	51	26.6	846	10 US-09-376-045-2	Sequence 2, Appli
12	51	26.6	908	10 US-09-376-045-6	Sequence 6, Appli
13	51	26.6	1053	9 US-10-217-700-6	Sequence 6, Appli
14	51	26.6	1054	10 US-09-376-045-4	Sequence 4, Appli
15	50	26.0	737	9 US-10-106-698-4893	Sequence 4893, Ap
16	50	26.0	738	9 US-10-235-521-1	Sequence 1, Appli
17	49	25.5	127	9 US-10-106-698-6472	Sequence 6472, Ap
18	49	25.5	1109	9 US-09-912-697-35	Sequence 35, Appl
19	48.5	25.3	125	10 US-09-864-761-38086	Sequence 38086, A

20	48.5	25.3	255	9 US-10-125-540-427	Sequence 427, App
21	48.5	25.3	255	10 US-09-764-870-427	Sequence 427, App
22	48.5	25.3	470	9 US-10-156-761-14879	Sequence 14879, A
23	48	25.0	354	9 US-10-050-704-115	Sequence 115, App
24	48	25.0	383	9 US-10-050-704-239	Sequence 239, App
25	48	25.0	511	9 US-10-156-761-11125	Sequence 11125, A
26	48	25.0	668	10 US-09-918-909-2	Sequence 2, Appli
27	48	25.0	1289	10 US-09-738-563-4	Sequence 4, Appli
28	47.5	24.7	245	8 US-08-424-550B-40	Sequence 40, Appli
29	47.5	24.7	270	9 US-09-738-636-5968	Sequence 5968, Ap
30	47.5	24.7	625	9 US-10-099-895-14	Sequence 34, Appli
31	47.5	24.7	1165	10 US-09-771-161A-239	Sequence 239, App
32	47.5	24.7	1061	9 US-10-156-761-14082	Sequence 14082, A
33	47.5	24.7	2597	10 US-09-905-129-2	Sequence 2, Appli
34	47.5	24.7	2597	10 US-09-905-129-10	Sequence 10, Appli
35	47.5	24.7	2597	10 US-09-905-129-13	Sequence 13, Appli
36	47.5	24.7	2597	10 US-09-991-630-2	Sequence 2, Appli
37	47.5	24.7	2597	10 US-09-991-630-10	Sequence 10, Appli
38	47.5	24.7	2597	10 US-09-991-630-13	Sequence 13, Appli
39	47.5	24.7	2972	8 US-08-424-550B-387	Sequence 387, App
40	47	24.5	373	9 US-09-978-295A-59	Sequence 59, Appl
41	47	24.5	373	9 US-09-992-598-503	Sequence 503, App
42	47	24.5	373	9 US-09-978-697-59	Sequence 59, Appl
43	47	24.5	373	9 US-09-978-192A-59	Sequence 59, Appl
44	47	24.5	373	9 US-09-989-293A-503	Sequence 503, App
45	47	24.5	373	9 US-09-989-735-503	Sequence 503, App

ALIGNMENTS

RESULT 1
US-09-984-130-43
; Sequence 43, Application US/09984130
; Publication No. US20030055231A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 12 Human Secreted Proteins
; FILE REFERENCE: PFA4892
; CURRENT APPLICATION NUMBER: US/09/984,130
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/243,792
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 09/836,353
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/198,407
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US99/25031
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 60/105,971
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 149
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-130-43
Query Match 30.7%; Score 59; DB 9; Length 1034;
Best Local Similarity 41.7%; Pred. No. 11;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
QY 12 GPGEFRTLRVQNLGCVVSGLI 35
Db 981 GIGPFSCIFRIQNLGFPFHGM 1004
RESULT 2
US-09-836-353A-43
; Sequence 43, Application US/09836353A
; Publication No. US20030129685A1
; GENERAL INFORMATION:

```
APPLICANT: Ni et al.
FILE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P1
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 43
LENGTH: 1034
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-43
```

```
Query Match          30.7% Score 59; DB 9; Length 1034;
Best Local Similarity 41.7% Pred. No. 11;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy      12 GPGPEFKTLRVQNLGCVVSGLI 35
Db      981 GIGPFSCIFRIQNLGLFPIHGIM 1004
```

```
RESULT 3
US-09-984-130-35
Sequence 35, Application US/09984130
Publication No. US20030055231A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P2
CURRENT APPLICATION NUMBER: US/09/984,130
CURRENT FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 60/243,792
PRIOR FILING DATE: 2000-10-30
PRIOR APPLICATION NUMBER: 09/836,353
PRIOR FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 149
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 35
LENGTH: 1189
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-130-35
```

```
Query Match          30.7% Score 59; DB 9; Length 1189;
Best Local Similarity 41.7% Pred. No. 13;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy      12 GPGPEFKTLRVQNLGCVVSGLI 35
Db      981 GIGPFSCIFRIQNLGLFPIHGIM 1004
```

```
RESULT 4
US-09-836-353A-35
Sequence 35, Application US/09836353A
Publication No. US20030129685A1
GENERAL INFORMATION:
APPLICANT: Ni et al.
TITLE OF INVENTION: 12 Human Secreted Proteins
FILE REFERENCE: PF489P1
```

```
CURRENT APPLICATION NUMBER: US/09/836,353A
CURRENT FILING DATE: 2001-04-18
PRIOR APPLICATION NUMBER: 60/198,407
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: PCT/US99/25031
PRIOR FILING DATE: 1999-10-27
PRIOR APPLICATION NUMBER: 60/105,971
PRIOR FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 147
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 35
LENGTH: 1189
TYPE: PRT
ORGANISM: Homo sapiens
US-09-836-353A-35
```

```
Query Match          30.7% Score 59; DB 9; Length 1189;
Best Local Similarity 41.7% Pred. No. 13;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
```

```
Qy      12 GPGPEFKTLRVQNLGCVVSGLI 35
Db      981 GIGPFSCIFRIQNLGLFPIHGIM 1004
```

```
RESULT 5
US-10-156-761-8464
Sequence 8464, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHITUKU
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8464
LENGTH: 3338
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8464
```

```
Query Match          29.9% Score 57.5; DB 9; Length 3338;
Best Local Similarity 56.5% Pred. No. 67;
Matches 13; Conservative 2; Mismatches 7; Indels 1; Gaps 1;
```

```
Qy      8 TLVPGPEFKTLRVQNLGCVV 30
Db      2091 TLVP-GPDPFSPARLRKGLGQV 2112
```

```
RESULT 6
US-09-866-050A-500
Sequence 500, Application US/09866050A
Publication No. US20030040477A1
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Ornst, Rene
APPLICANT: Murison, James G.
APPLICANT: Kumble, Krishanand D.
TITLE OF INVENTION: Compositions Isolated From Skin Cells
```


;; TITLE OF INVENTION: and Methods for Their Use
;; FILE REFERENCE: 11000.1011c4U
;; CURRENT APPLICATION NUMBER: US/09/866,050A
;; CURRENT FILING DATE: 2001-05-24
;; NUMBER OF SEQ ID NOS: 725
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 500
;; LENGTH: 545
;; TYPE: PRT
;; ORGANISM: Mouse
US-09-866-050A-500

Query Match 29.2%; Score 56; DB 9; Length 545;
Best Local Similarity 41.7%; Pred. No. 14;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 12 GGPPEFKTLRVONLGCYVSGLI 35
DB 338 GIGPPEFNCVFKVQNLGFFPIHGV 361

RESULT 7
US-09-866-050A-624
;; Sequence 624, Application US/09866050A
;; Publication No. US20030040471A1
;; GENERAL INFORMATION:
;; APPLICANT: Watson, James D.
;; APPLICANT: Strachan, Lorna
;; APPLICANT: Sleeman, Matthew
;; APPLICANT: Onrust, Rene
;; APPLICANT: Murison, James G.
;; APPLICANT: Kumble, Krishanand D.
;; TITLE OF INVENTION: Compositions Isolated From Skin Cells
;; FILE REFERENCE: 11000.1011c4U
;; CURRENT APPLICATION NUMBER: US/09/866,050A
;; CURRENT FILING DATE: 2001-05-24
;; NUMBER OF SEQ ID NOS: 725
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 624
;; LENGTH: 688
;; TYPE: PRT
;; ORGANISM: Rat
US-09-866-050A-624

Query Match 29.2%; Score 56; DB 9; Length 688;
Best Local Similarity 41.7%; Pred. No. 18;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 12 GGPPEFKTLRVONLGCYVSGLI 35
DB 481 GIGPPEFNCVFKVQNLGFFPIHGV 504

RESULT 8
US-09-866-050A-501
;; Sequence 501, Application US/09866050A
;; Publication No. US20030040471A1
;; GENERAL INFORMATION:
;; APPLICANT: Watson, James D.
;; APPLICANT: Strachan, Lorna
;; APPLICANT: Sleeman, Matthew
;; APPLICANT: Onrust, Rene
;; APPLICANT: Murison, James G.
;; APPLICANT: Kumble, Krishanand D.
;; TITLE OF INVENTION: Compositions Isolated From Skin Cells
;; FILE REFERENCE: 11000.1011c4U
;; CURRENT APPLICATION NUMBER: US/09/866,050A
;; CURRENT FILING DATE: 2001-05-24
;; NUMBER OF SEQ ID NOS: 725
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 501

;; LENGTH: 696
;; TYPE: PRT
;; ORGANISM: Rat
US-09-866-050A-501

Query Match 29.2%; Score 56; DB 9; Length 696;
Best Local Similarity 41.7%; Pred. No. 18;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

OY 12 GGPPEFKTLRVONLGCYVSGLI 35
DB 481 GIGPPEFNCVFKVQNLGFFPIHGV 504

RESULT 9
US-10-156-761-13191
;; Sequence 13191, Application US/10156761
;; Publication No. US20030119018A1
;; GENERAL INFORMATION:
;; APPLICANT: OMURA, SATOSHI
;; APPLICANT: IKEDA, HARUO
;; APPLICANT: HORIKAWA, HIROSHI
;; APPLICANT: SHIBA, TADAYOSHI
;; APPLICANT: SAKAKI, YOSHIYUKI
;; APPLICANT: HATTORI, MASAHIRA
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-262
;; CURRENT APPLICATION NUMBER: US/10/156,761
;; CURRENT FILING DATE: 2002-05-29
;; PRIOR APPLICATION NUMBER: JP 2001-204089
;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO 13191
;; LENGTH: 308
;; TYPE: PRT
;; ORGANISM: Streptomyces avermitilis
US-10-156-761-13191

Query Match 27.1%; Score 52; DB 9; Length 308;
Best Local Similarity 38.7%; Pred. No. 26;
Matches 12; Conservative 6; Mismatches 9; Indels 4; Gaps 1;

OY 5 PYGTLPVGPPEFKTLRVONLGCYVSGLI 35
DB 105 PYGTADG---FETQFGVNHIGHFALTGL 131

RESULT 10
US-09-753-831-2
;; Sequence 2, Application US/09753831
;; Patent No. US20020137683A1
;; GENERAL INFORMATION:
;; APPLICANT: Hogan, Kevin T.
;; APPLICANT: ROSS, Mark W.
;; TITLE OF INVENTION: C-SKI Oncogene-Derived Peptides for Prevention,
;; Treatment and Diagnosis of Cancer
;; FILE REFERENCE: 26747-27
;; CURRENT APPLICATION NUMBER: US/09/753,831
;; CURRENT FILING DATE: 2001-01-03
;; PRIOR APPLICATION NUMBER: U.S. 60/174296
;; PRIOR FILING DATE: 2000-01-03
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 2
;; LENGTH: 728
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-753-831-2

Query Match 27.1%; Score 52; DB 10; Length 728;

Best Local Similarity 35.3%; Pred. No. 69;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

Qy 10 PVGPGEF-----KTLRVQNLGCVVSG 33
Db 84 PVLPGFFMPDSRSTERCETLEGETISCFVVG 117

RESULT 11

US-09-376-045-2
; Sequence 2, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-376-045-2

Query Match 26.6%; Score 51; DB 10; Length 846;
Best Local Similarity 35.5%; Pred. No. 1.1e+02;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFCKTLRVQNLGCVV 31
Db 704 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 732

RESULT 12

US-09-376-045-6
; Sequence 6, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-376-045-6

Query Match 26.6%; Score 51; DB 10; Length 908;
Best Local Similarity 35.5%; Pred. No. 1.2e+02;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFCKTLRVQNLGCVV 31
Db 726 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 13

US-10-217-700-6
; Sequence 6, Application US/10217700
; Publication No. US20030070191A1
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holdaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE

FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/10/217,700
; CURRENT FILING DATE: 2002-08-12
; EARLIER APPLICATION NUMBER: 09/394,272
; EARLIER FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1053
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-10-217-700-6

Query Match 26.6%; Score 51; DB 9; Length 1053;
Best Local Similarity 35.5%; Pred. No. 1.5e+02;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFCKTLRVQNLGCVV 31
Db 911 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 939

RESULT 14

US-09-376-045-4
; Sequence 4, Application US/09376045
; Patent No. US20020019998A1
; GENERAL INFORMATION:
; APPLICANT: Aventis CropScience GmbH
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE PREPARATION OF
; TITLE OF INVENTION: PLANTS WITH CHANGED SUCROSE CONCENTRATION
; FILE REFERENCE: 514413-3772
; CURRENT APPLICATION NUMBER: US/09/376,045
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1054
; TYPE: PRT
; ORGANISM: Solanum tuberosum
US-09-376-045-4

Query Match 26.6%; Score 51; DB 10; Length 1054;
Best Local Similarity 35.5%; Pred. No. 1.5e+02;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

Qy 1 YEHPYGLPVGPGEFCKTLRVQNLGCVV 31
Db 912 FKVCKPGTVP--PSKELRKVMRIQALRCHAV 940

RESULT 15

US-10-106-698-4893
; Sequence 4893, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: P000501
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4893
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 1.37834 Seconds
(without alignments)
747.132 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192
Sequence: 1 YEHPYGTLPVGPDPFRTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents, AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	ID	Description
1	52	27.1	750	4	US-08-202-841A-2
2	51	26.6	846	1	US-08-356-354-2
3	51	26.6	846	1	US-08-778-656-2
4	51	26.6	890	1	US-08-145-006C-5
5	51	26.6	890	5	PCT-US94-00545-5
6	51	26.6	908	1	US-08-356-354-6
7	51	26.6	908	2	US-08-778-656-6
8	51	26.6	1054	1	US-08-356-354-4
9	51	26.6	1054	2	US-08-778-656-4
10	50	26.0	738	4	US-08-989-385-1
11	48	25.0	668	4	US-09-697-367-2
12	48	25.0	1289	1	US-07-876-280-4
13	48	25.0	1289	1	US-07-675-772-4
14	48	25.0	1289	1	US-08-063-170-4
15	48	25.0	1289	1	US-08-158-232-4
16	48	25.0	1289	1	US-08-304-626-4
17	48	25.0	1289	1	US-08-316-301A-4
18	48	25.0	1289	2	US-08-611-928-4
19	48	25.0	1289	3	US-09-173-891-4
20	48	25.0	1289	4	US-09-076-137-4
21	48	25.0	1289	5	PCT-US92-03624-4
22	48	25.0	1289	5	5281530-3
23	48	25.0	1289	6	5426049-4
24	47.5	24.7	245	4	US-08-469-260A-40
25	47.5	24.7	1065	4	US-09-412-545-2
26	47.5	24.7	2972	4	US-08-469-260A-387
27	47	24.5	1057	4	US-08-853-948B-2

28	47	24.5	1057	4	US-09-697-367-23	Sequence 23, Appl
29	46	24.0	33	1	US-08-257-528B-84	Sequence 84, Appl
30	46	24.0	33	1	US-08-460-602A-84	Sequence 84, Appl
31	46	24.0	33	1	US-08-463-966A-84	Sequence 84, Appl
32	46	24.0	33	1	US-08-465-217A-84	Sequence 84, Appl
33	46	24.0	33	2	US-08-464-329A-84	Sequence 84, Appl
34	46	24.0	33	2	US-08-462-507A-84	Sequence 84, Appl
35	46	24.0	33	2	US-08-467-881A-84	Sequence 84, Appl
36	46	24.0	34	4	US-09-141-833-11	Sequence 11, Appl
37	46	24.0	1045	4	US-08-553-436A-6	Sequence 6, Appl
38	46	24.0	1193	4	US-09-227-725A-4	Sequence 4, Appl
39	45.5	23.7	33	1	US-08-257-528B-83	Sequence 83, Appl
40	45.5	23.7	33	1	US-08-460-602A-83	Sequence 83, Appl
41	45.5	23.7	33	1	US-08-463-966A-83	Sequence 83, Appl
42	45.5	23.7	33	1	US-08-465-217A-83	Sequence 83, Appl
43	45.5	23.7	33	2	US-08-464-329A-83	Sequence 83, Appl
44	45.5	23.7	33	2	US-08-462-507A-83	Sequence 83, Appl
45	45.5	23.7	33	2	US-08-467-881A-83	Sequence 83, Appl

ALIGNMENTS

RESULT 1
US-08-202-841A-2
Sequence 2, Application US/08202841A
Patent No. 6218596
GENERAL INFORMATION:
APPLICANT: Hughes, Stephen H.
APPLICANT: Sutlive, Pirmod
APPLICANT: Pursel, Vernon
TITLE OF INVENTION: Enhancement of Musculature in Animals
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,841A
FILING DATE: 23-FEB-1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620,415
FILING DATE: 03-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,449
FILING DATE: 02-JUL-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/373,864
FILING DATE: 30-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: Alicea, Hector A.
REGISTRATION NUMBER: 40,891
REFERENCE/DOCKET NUMBER: 015280-170300US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 750 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Modified-site

LOCATION: 373
OTHER INFORMATION: /product= "OTHER"
OTHER INFORMATION: /note= "Xaa = Trp in c-ski";
OTHER INFORMATION: Xaa = Arg in v-ski"
US-08-202-841A-2

Query Match 27.1%; Score 52; DB 4; Length 750;
Best Local Similarity 35.3%; Pred. No. 21;
Matches 12; Conservative 4; Mismatches 8; Indels 10; Gaps 1;

QY 10 PVGPGPEF-----KTLRVQNLGCVVSG 33
DB 67 PVMPGPFMPSPDRSTERCTILEGTISCVVGS 100

RESULT 2
US-08-356-354-2

Sequence 2, Application US/08356354
Patent No. 5767365

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

NUMBER OF SEQUENCES: 6 PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

STREET: 1180 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,354

FILING DATE: 20-DEC-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-356-354-2

Query Match 26.6%; Score 51; DB 1; Length 846;

Best Local Similarity 35.5%; Pred. No. 33;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTLRVQNLGCVV 31

DB 704 FKVKCKGTVP--PSKELRKVMRIQLRCHAV 732

RESULT 3

US-08-778-656-2

Sequence 2, Application US/08778656

Patent No. 5976869

GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe

TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE

NUMBER OF SEQUENCES: 6 PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

STREET: 1180 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10036-8403

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/778,656

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/356,354

FILING DATE: 20-DEC-1994

APPLICATION NUMBER: US PCT/EP93/01605

FILING DATE: 22-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P42 20 758.4

FILING DATE: 24-JUN-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/951-105

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-778-656-2

Query Match 26.6%; Score 51; DB 2; Length 846;

Best Local Similarity 35.5%; Pred. No. 33;

Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTLRVQNLGCVV 31

DB 704 FKVKCKGTVP--PSKELRKVMRIQLRCHAV 732

RESULT 4

US-08-145-006C-5

Sequence 5, Application US/08145006C

Patent No. 5656452

GENERAL INFORMATION:

APPLICANT: Rao, Anjana

APPLICANT: Hogan, Patrick Gerald

APPLICANT: McCaffrey, Patricia

APPLICANT: Jain, Jugnu

TITLE OF INVENTION: NF-ATp, A T LYMPHOCYTE

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/145,006C
FILING DATE: October 29, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/017,052
FILING DATE: February 11, 1993
APPLICATION NUMBER: 08/006,067
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Frazer, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 04590/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 890
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-145-006C-5

Query Match 26.6%; Score 51; DB 1; Length 890;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPYGLPYGPGP 15
|||
Db 805 HPLGTCVLPGP 816

RESULT 5
PCT-US94-00545-5
Sequence 5, Application PC/TUS9400545
GENERAL INFORMATION:
APPLICANT: Rao, Anjana
APPLICANT: Hogan, Patrick Gerald
APPLICANT: McCaffrey, Patricia
APPLICANT: Jain, Jyugnu
TITLE OF INVENTION: NF-ATP, A T LYMPHOCYTE
TITLE OF INVENTION: DNA-BINDING PROTEIN
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00545
FILING DATE: 18-JAN-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/145,006
FILING DATE: October 29, 1993
APPLICATION NUMBER: 08/017,052
FILING DATE: February 11, 1993

APPLICATION NUMBER: 08/006,067
FILING DATE: January 15, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Frazer, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 04590/007001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 890
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US94-00545-5

Query Match 26.6%; Score 51; DB 5; Length 890;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 HPYGLPYGPGP 15
|||
Db 805 HPLGTCVLPGP 816

RESULT 6
US-08-356-354-6
Sequence 6, Application US/08356354
Patent No. 5767365
GENERAL INFORMATION:
APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
TITLE OF INVENTION: PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oestrolekn, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,354
FILING DATE: 20-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/EP93/01605
FILING DATE: 22-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42 20 758.4
FILING DATE: 24-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-356-354-6

Query Match 26.6%; Score 51; DB 1; Length 908;
Best Local Similarity 35.5%; Pred. No. 36;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31
DB 726 FKVCCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 7

US-08-778-656-6
Sequence 6, Application US/08778656
Patent No. 5976869
GENERAL INFORMATION:
APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,656
FILING DATE:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-778-656-6

Query Match 26.6%; Score 51; DB 2; Length 908;
Best Local Similarity 35.5%; Pred. No. 36;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31
DB 726 FKVCCKPGTVP--PSKELRKVMRIQALRCHAV 754

RESULT 8

US-08-356-354-4
Sequence 4, Application US/08356354
Patent No. 5767365
GENERAL INFORMATION:

APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US PCT/EP93/01605
FILING DATE: 22-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42 20 758.4
FILING DATE: 24-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-105
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1054 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-356-354-4

Query Match 26.6%; Score 51; DB 1; Length 1054;
Best Local Similarity 35.5%; Pred. No. 43;
Matches 11; Conservative 7; Mismatches 11; Indels 2; Gaps 1;

QY 1 YEYHPYGLPVGPGPEFKTTLRVONIGCYV 31
DB 912 FKVCCKPGTVP--PSKELRKVMRIQALRCHAV 940

RESULT 9

US-08-778-656-4
Sequence 4, Application US/08778656
Patent No. 5976869
GENERAL INFORMATION:
APPLICANT: SONNEWALD, Uwe
TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
PREPARATION OF PLANTS WITH CHANGED SUCROSE CONCENTRATION
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/876,280
FILING DATE: 19920430
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
CLONE: E. coli NMS22 (pMYC 1628) NRRL B-18652
US-07-876-280-4

Query Match 25.0%; Score 48; DB 1; Length 1289;
Best Local Similarity 38.7%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLPU-GPGPERKTLRVQ--NLGCVVSGL 34
Db 743 TPIPGSGKDFNTLEIDIVSIDIFVSGL 773

RESULT 13
US-07-675-772-4
Sequence 4, Application US/07675772
Patent No. 5262399
GENERAL INFORMATION:
APPLICANT: Hickie, Leslie A.
APPLICANT: Sick, August J.
APPLICANT: Schwab, George E.
APPLICANT: Narva, Kenneth E.
APPLICANT: Payne, Jewel M.
TITLE OF INVENTION: No. 5262399el Compositions and Methods for the Control of
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROMAN SALIWANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/675,772
FILING DATE: 19910327
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, ROMAN
REGISTRATION NUMBER: 21,023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS17
IMMEDIATE SOURCE:
LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF KENNETH NARVA
CLONE: 17B
US-07-675-772-4

Query Match 25.0%; Score 48; DB 1; Length 1289;
Best Local Similarity 38.7%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLPU-GPGPERKTLRVQ--NLGCVVSGL 34
Db 743 TPIPGSGKDFNTLEIDIVSIDIFVSGL 773

RESULT 14
US-08-063-170-4
Sequence 4, Application US/08063170
Patent No. 5350576
GENERAL INFORMATION:
APPLICANT: Kim, Leo
APPLICANT: Schwab, George E.
TITLE OF INVENTION: Compositions and Methods for Inducing an Immune
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID R. SALIWANCHIK
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/063,170
FILING DATE: 19930517
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,141
FILING DATE: 30-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/759,248
FILING DATE: 13-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: SALIWANCHIK, DAVID R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 103.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17b
US-08-063-170-4

Query Match 25.0%; Score 48; DB 1; Length 1289;
Best Local Similarity 38.7%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLVP-GPGPERFTLRVQ--NLGCVVSGL 34
Db 743 TIPPGSGKDFNTLEIODIVSIDIFVSGSL 773

RESULT 15
US-08-158-232-4
Sequence 4, Application US/08158232
Patent No. 5596071
GENERAL INFORMATION:

APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Poncerada, Luis
APPLICANT: Schepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
NUMBER OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,232
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
STRAIN: PS17
INDIVIDUAL ISOLATE: PS17b
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC1628) NRRL B-18652
US-08-158-232-4

Query Match 25.0%; Score 48; DB 1; Length 1289;
Best Local Similarity 38.7%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 8; Indels 4; Gaps 2;

Qy 8 TLVP-GPGPERFTLRVQ--NLGCVVSGL 34
Db 743 TIPPGSGKDFNTLEIODIVSIDIFVSGSL 773

Search completed: July 16, 2003, 07:57:48
Job time : 2.37834 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:21:09 ; Search time 39 Seconds
(without alignments)
86,274 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192
Sequence: 1 YEHPEYGTLPVGPGEPEFKTLRVNLCGYVWSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 7751

Minimum DB seq length: 0
Maximum DB seq length: 35

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	19.3	35	2	PC2297
2	36	18.8	35	2	S05414
3	35.5	18.5	33	2	S13863
4	35	18.2	33	2	D61563
5	35	18.2	33	2	A54257
6	34.5	18.0	35	2	B36912
7	34	17.7	35	2	PC2295
8	33	17.2	10	2	H60588
9	33	17.2	30	2	T69492
10	32	16.7	22	2	P00070
11	32	16.7	30	2	S74112
12	32	16.7	35	2	PC2296
13	32	16.7	35	2	B41161
14	31.5	16.4	35	2	AD2525
15	31	16.1	19	2	S68394
16	31	16.1	20	2	A36016
17	31	16.1	24	2	T42441
18	31	16.1	24	2	B30609
19	31	16.1	28	2	A31859
20	31	16.1	28	2	JN0366
21	31	16.1	34	2	CJ1514
22	30.5	15.9	35	2	PC2294
23	30.5	15.9	35	2	S18264
24	30	15.6	18	2	C32537
25	30	15.6	20	2	PL0161
26	30	15.6	20	2	D32537
27	30	15.6	27	2	C54257
28	30	15.6	28	2	S41774
29	30	15.6	28	2	I48349

30	30	15.6	33	2	A95047	hypothetical prote
31	30	15.6	34	2	H95019	hypothetical prote
32	30	15.6	35	2	PC2293	V3 domain peptide
33	29.5	15.4	27	2	A24487	aldose 1-epimerase
34	29	15.1	15	2	S08282	cytochrome P450K-2
35	29	15.1	22	2	A39269	LX-1 tumor antigen
36	29	15.1	27	2	S64717	formin binding pro
37	29	15.1	30	2	A34874	transforming prote
38	29	15.1	35	2	PC2298	V3 domain peptide
39	28.5	14.8	33	2	A60507	pepsin I (EC 3.4.2
40	28	14.6	19	2	S69153	Neb-collagenase
41	28	14.6	20	2	S65399	immunodeficiency v
42	28	14.6	20	2	P00033	aldose 1-epimerase
43	28	14.6	22	2	H30608	ig kappa chain V-I
44	28	14.6	25	2	S36378	ig heavy chain V r
45	28	14.6	26	2	G30608	ig kappa chain V-I

ALIGNMENTS

RESULT 1
PC2297
V3 domain peptide P4611 - human immunodeficiency virus type 1 (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 21-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 07-May-1999
C:Accession: PC2297
R.Sherefa, K.; Soenherborg, A.; Steinberg, J.; Saelberg, M.
Biochem. Biophys. Res. Commun. 205:1658-1664, 1994
A:Title: Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping a
A:Reference number: PC2291; MUID:95110306; PMID:7811250
A:Accession: PC2297
A:Molecule type: protein
A:Residues: 1-35 <SHE>
C:Superfamily: type E retrovirus env polyprotein

Query Match 19.3% Score 37; DB 2; Length 35;
Best Local Similarity 46.2% Pred. No. 1.7e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLVPVGPGEPEFKTT 20
DB 11 SIPIGGRALYTT 23

RESULT 2
S05414
steryl-sulfatase (EC 3.1.6.2), microsomal - rat (fragment)
N:Alternate names: arylsulfatase C
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 28-Apr-1993
C:Accession: S05414
R.Kawano, U.I.; Kotani, T.; Ohteki, S.; Minamino, N.; Matsuo, H.; Oinuma, T.; Aikawa, E.
Biochim. Biophys. Acta 997, 199-205, 1989
A:Title: Characterization of rat and human steroid sulfatases.
A:Reference number: S05414; MUID:89352671; PMID:2765556
A:Accession: S05414
A:Molecule type: protein
A:Residues: 1-26 <KAW>
C:Keywords: sulfuric ester hydrolase

Query Match 18.8% Score 36; DB 2; Length 26;
Best Local Similarity 40.9% Pred. No. 1.7e+02;
Matches 9; Conservative 2; Mismatches 7; Indels 4; Gaps 1;

QY 12 GPGPEF---KTLRVNLCGY 29
DB 4 GPGPEFLIMADDLGIGLGY 25

RESULT 3
S13863
dihydrolipoamide dehydrogenase (EC 1.8.1.4) - Trypanosoma cruzi (fragment)

C:Species: Trypanosoma cruzi
C:Date: 18-Feb-1994 #sequence_revision 06-Sep-1996 #text_change 23-May-1997

C:Accession: S13863

R:Locher, H.; Krauth-Siegel, R.L.
Eur. J. Biochem. 194, 863-869, 1990

A:Title: Purification and characterization of lipamide dehydrogenase from Trypanosoma
A:Reference number: S13863; MUID:91099369; PMID:2269305

A:Accession: S13863

A:Molecule type: protein

A:Residues: 1-33 <LOH>

C:Superfamily: dihydrolipamide dehydrogenase; dihydrolipamide dehydrogenase homology
C:Keywords: FAD; flavoprotein; homodimer; lipamide; NAD; oxidoreductase; redox-active
F4-32/Region: beta-alpha-beta FAD nucleotide-binding fold

Query Match 18.5%; Score 35.5; DB 2; Length 33;
Best Local Similarity 31.0%; Pred. No. 2.6e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 11; Gaps 1;

QY 4 HPYGLPVGPGEFKTLRVQNLGCVVS 32
DB 1 NPYDVVIGGP-----GTVAS 18

RESULT 4

D61563

neurophysin 1 - chicken (fragments)

C:Species: Gallus gallus (chicken)

C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 17-Mar-1999

C:Accession: D61563

R:Levy, B.; Michel, G.; Chauvet, J.; Chauvet, M.T.; Acher, R.

Biochem. Rep. 7, 631-636, 1987

A:Title: Gene conversion in avian mesocortin and vasotocin genes: a recurrent mechanism

A:Reference number: A61563; MUID:88108074; PMID:3427215

A:Accession: D61563

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-33 <LEV>

C:Superfamily: oxytocin-neurophysin

C:Keywords: hormone

Query Match 18.2%; Score 35; DB 2; Length 33;
Best Local Similarity 37.0%; Pred. No. 3.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 8; Gaps 2;

QY 9 LPVGP-----GPEFKTLRVQNLGCVV 30
DB 3 LFGPRNKCFCFPRNCG---EELGCTV 26

RESULT 5

A54257

deoxynucleoside kinase complex I S-component - Lactobacillus acidophilus (fragment)

C:Species: Lactobacillus acidophilus

C:Date: 06-Oct-1994 #sequence_revision 18-Nov-1994 #text_change 18-Sep-1998

C:Accession: A54257

R:Iikeda, S.; Ma, G.T.; Ives, D.H.

Biochemistry 33, 5328-5334, 1994

A:Title: Heterodimeric deoxynucleoside kinases of Lactobacillus acidophilus R-26: functi

A:Reference number: A54257; MUID:94227067; PMID:8172906

A:Accession: A54257

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-35 <IKE>

A:Experimental source: R-26

A:Note: sequence extracted from NCBI backbone (NCBI:146748)

C:Superfamily: Lactobacillus acidophilus deoxyadenosine kinase

Query Match 18.2%; Score 35; DB 2; Length 35;
Best Local Similarity 38.9%; Pred. No. 3.3e+02;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 10 PVGPGEFKTLRVQNLG 27

DB 7 PIGAGSSULTSLAEXLG 24

RESULT 6

B36912

hypothetical protein 2 baIG-region [imported] - Eubacterium sp. (fragment)

C:Species: Eubacterium sp.

C:Date: 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 20-Jun-2000

C:Accession: B36912

R:Franklund, C.V.; Baron, S.F.; Hylemon, P.B.

J. Bacteriol. 175, 3002-3012, 1993

A:Title: Characterization of the baIG gene encoding a bile acid-inducible NADH:flavin ox

A:Reference number: A36912; MUID:93259945; PMID:8491719

A:Accession: B36912

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-35 <FRA>

A:Experimental source: VPI 12708

A:Note: sequence extracted from NCBI backbone (NCBI:131908)

Query Match 18.0%; Score 34.5; DB 2; Length 35;
Best Local Similarity 56.2%; Pred. No. 3.9e+02;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 20 TLRVQNLGCVVSGLI 35
DB 7 TLTVLELCIV-GII 21

RESULT 7

PC2295

V3 domain peptide P4317 - human immunodeficiency virus type 1 (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 21-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 07-May-1999

C:Accession: PC2295

R:Sherefa, K.; Soenarborg, A.; Steinberg, J.; Saelberg, M.

Biochem. Biophys. Res. Commun. 205, 1658-1664, 1994

A:Title: Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping a

A:Reference number: PC2291; MUID:95110306; PMID:7811250

A:Accession: PC2295

A:Molecule type: protein

A:Residues: 1-35 <SHE>

C:Superfamily: type E retrovirus env polyprotein

Query Match 17.7%; Score 34; DB 2; Length 35;
Best Local Similarity 46.2%; Pred. No. 4.6e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLPVGPGEFKTT 20
DB 11 SIHIGPGRFYTT 23

RESULT 8

H60588

sperm-activating peptide (Thr-2,5, Leu-3, Pro-4, Ser-7,10 SAP-I) - slate-pencil urchin (f

C:Species: Heterocentrotus mamillatus

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 18-Aug-2000

C:Accession: H60588

R:Yoshino, K.I.; Kajitara, H.; Nomura, K.; Takao, T.; Shimomishi, Y.; Kurita, M.; Yamaguchi

Comp. Biochem. Physiol. B 94, 739-751, 1989

A:Title: A halogenated amino acid-containing sperm activating peptide and its related pe

ctus nudus, Echionometra mathaei and Heterocentrotus mamillatus.

A:Reference number: A60527

A:Accession: H60588

A:Molecule type: protein

A:Residues: 1-10 <YOS>

C:Superfamily: unassigned animal peptides

Query Match 17.2%; Score 33; DB 2; Length 10;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 GTLPVGP 14
|||
Db 1 GTLPVGP 8

RESULT 9

169492
gene aeg-46.5 protein - Escherichia coli
C:Species: Escherichia coli
C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 08-Oct-1999
C:Accession: 169492
R:Choe, M.; Reznikoff, W.S.
J. Bacteriol. 175, 1165-1172, 1993
A:Title: Identification of the regulatory sequence of an aerobically expressed locus aeg-4
A:Reference number: 154984; MUID:93163046; PMID:8432709
A:Accession: 169492
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-30 <RES>
A:Cross-references: GB:S54754; NID:9265388; PID:AA825330.1; PID:9265389

Query Match

17.7%; Score 32; DB 2; Length 30;
Best Local Similarity 31.4%; Pred. No. 5.4e+02;
Matches 11; Conservative 2; Mismatches 4; Indels 18; Gaps 2;

QY 3 VHPVG-----TLVGPGEPEKTTLRVQNL 26
|||
Db 1 MHPVGASLVAGAKPVTVSVRGP-----VMNL 28

RESULT 10

P00070
T-cell receptor beta chain (BTB15) - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Feb-1995
C:Accession: P00070
R:Tanaka, A.; Ishiguro, N.; Shinagawa, M.
submitted to JIPID, May 1990
A:Description: Sequence analysis of bovine T-cell receptor beta chain genes.
A:Reference number: J00472
A:Accession: P00070
A:Molecule type: mRNA
A:Residues: 1-22 <TRAN>
A:Experimental source: T cell
C:Genetics:
A:Gene: BTB15
C:Keywords: receptor

Query Match 16.7%; Score 32; DB 2; Length 22;
Best Local Similarity 66.7%; Pred. No. 5.3e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 YGTLFVGP 14
|||
Db 8 YGTLFVGP 16

RESULT 11

S74112
proline-rich antibacterial protein - green crab (fragment)
C:Species: Carcinus maenas (green crab, common shore crab)
C:Date: 11-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 17-Apr-1998
C:Accession: S74112
R:Schnapf, D.; Kemp, G.D.; Smith, V.J.
Eur. J. Biochem. 240, 532-539, 1996
A:Title: Purification and characterization of a proline-rich antibacterial peptide with
A:Accession: S74112
A:Reference number: S74112; MUID:97008941; PMID:8856051
A:Molecule type: protein
A:Residues: 1-30 <SCH>
A:Experimental source: haemocytes
C:Keywords: antibacterial

Query Match 16.7%; Score 32; DB 2; Length 30;
Best Local Similarity 45.5%; Pred. No. 7.4e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PYGTLVGP 15
|||
Db 8 PPRPPIGPR 18

RESULT 12

PC2296
V3 domain peptide P4346 - human immunodeficiency virus type 1 (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 21-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 07-May-1999
C:Accession: PC2296
R:Shereff, K.; Soenenborg, A.; Steinberg, J.; Saelberg, M.
Biochem. Biophys. Res. Commun. 205, 1658-1664, 1994
A:Title: Rapid grouping of HIV-1 infection in subtypes A to E by V3 peptide serotyping a
A:Reference number: PC2291; MUID:95110306; PMID:7811250
A:Accession: PC2296
A:Molecule type: protein
A:Residues: 1-35 <SHE>
C:Superfamily: type E retrovirus env polyprotein

Query Match 16.7%; Score 32; DB 2; Length 35;
Best Local Similarity 46.2%; Pred. No. 8.8e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLPVGPGEPEKTT 20
|||
Db 11 SHMGPRAFYTT 23

RESULT 13

B41161
2XK antigen PEB2 - Campylobacter jejuni (fragment)
C:Species: Campylobacter jejuni
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 23-Mar-1993
C:Accession: B41161
R:Pei, Z.; Ellison III, R.T.; Blaser, M.J.
J. Biol. Chem. 266, 16363-16369, 1991
A:Title: Identification, purification, and characterization of major antigenic proteins
A:Reference number: A41161; MUID:91358413; PMID:1885571
A:Accession: B41161
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-35 <PEI>

Query Match 16.7%; Score 32; DB 2; Length 35;
Best Local Similarity 47.1%; Pred. No. 8.8e+02;
Matches 8; Conservative 1; Mismatches 6; Indels 2; Gaps 1;

QY 2 EVHPYGTLPVGPGEPEK 18
|||
Db 1 EILVYG--PGGPAPVLK 15

RESULT 14

AD2525
hypothetical protein asr7380 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2525
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Itiguchi
Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2525
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-35 <KUR>

A_Cross-references: GB:BA000020; PION:BAH7138.1; PID:gl1134579; GSPDB:GN00180
A_Experimental source: strain PCC 7120
C_Genetics:
A_Gene: asr7380
A_Genome: Plasmid

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:15:09 ; Search time 22 Seconds
(without alignments)
65.985 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192
Sequence: 1 YEYHPYGTLPVGPSPFKTLRVNLTGCVVSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 2327

Minimum DB seq length: 0
Maximum DB seq length: 35

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	34	17.7	28	1	C1QC_RAT
2	32	16.7	30	1	AP65_CARMA
3	31	16.1	34	1	HEMO_CHICK
4	28	14.6	19	1	COOT_SARBU
5	28	14.6	28	1	PA22_MICNI
6	27	14.1	17	1	NID4M_TIRIU
7	27	14.1	22	1	RI18_HALME
8	27	14.1	23	1	RI18_HALVO
9	27	14.1	26	1	RI18_HALNA
10	27	14.1	27	1	PA21_MICNI
11	27	14.1	28	1	PA23_MICNI
12	27	14.1	30	1	PAC2_PSEPO
13	27	14.1	30	1	RI18_HALCU
14	27	14.1	31	1	CY06_VI00D
15	27	14.1	31	1	CY0B_VI00D
16	27	14.1	32	1	NEUB_PIG
17	27	14.1	33	1	Y656_TREPA
18	26	13.5	18	1	CPAX_BOVIN
19	26	13.5	30	1	NJ5W_FISOC
20	25	13.0	13	1	HMD_METWO
21	25	13.0	19	1	BPT1_BOTUA
22	25	13.0	15	1	MCRA_METTE
23	24.5	12.8	15	1	R86_PHYPA
24	24.5	12.8	30	1	FIBR_PANIN
25	24.5	12.8	32	1	CAL_ANGUA
26	24	12.5	7	1	UN06_PINPS
27	24	12.5	15	1	UC08_MAIZE
28	24	12.5	18	1	AHD2_TETPY
29	24	12.5	24	1	CAMT_PINPS
30	24	12.5	24	1	RAN_XENLA
31	24	12.5	26	1	CATG_RAT
32	24	12.5	31	1	ER29_BOVIN
33	24	12.5	32	1	PA22_AGRHP

34	24	12.5	35	1	CECB_ANTPE
35	24	12.5	35	1	SPRC_PIG
36	23.5	12.2	30	1	PLMS_SQUAC
37	23	12.0	16	1	HBD_GLOPA
38	23	12.0	16	1	LE05_BIOGL
39	23	12.0	27	1	PSAF_SYNTU
40	23	12.0	29	1	PSAF_SYNP6
41	23	12.0	31	1	CIRB_CHAPA
42	23	12.0	31	1	PETL_CHLVU
43	23	12.0	33	1	PAP1_PARV
44	23	12.0	33	1	PAP2_PARV
45	23	12.0	33	1	PAP3_PARV

ALIGNMENTS

RESULT 1	
CIQC_RAT	STANDARD; PRT; 28 AA.
AC P31722;	
DT 01-JUL-1993 (Rel. 26, Created)	
DT 01-JUL-1993 (Rel. 26, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Complement C1q subcomponent, C chain (Fragment).	
GN C1QG OR C1QC.	
OS Rattus norvegicus (Rat).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX NCBI_Taxid=10116;	
RN [1]	
RP MEDLINE:93218657; PubMed:8464426;	
RA Wing M.G., Seilly D.J., Bridgman R.A.;	
RT "Rapid isolation and biochemical characterization of rat C1 and C1q."	
RL Mol. Immunol. 30:433-440(1993).	
CC -1- FUNCTION: C1Q ASSOCIATES WITH THE PROENZYMES C1R AND C1S TO YIELD	
CC C1, THE FIRST COMPONENT OF THE SERUM COMPLEMENT SYSTEM. THE	
CC COLLAGEN-LIKE REGIONS OF C1Q INTERACT WITH THE C1(2+)-DEPENDENT	
CC C1R(2)C1S(2) PROENZYME COMPLEX, AND EFFICIENT ACTIVATION OF C1	
CC TAKES PLACE ON INTERACTION OF THE GLOBULAR HEADS OF C1Q WITH THE	
CC FC REGIONS OF IGG OR IGM ANTIBODY PRESENT IN IMMUNE COMPLEXES.	
CC -1- SUBUNIT: C1 IS A CALCIUM-DEPENDENT TRIMOLECULAR COMPLEX OF C1Q, R	
CC AND S IN THE MOLAR RATION OF 1:2:2. C1Q SUBCOMPONENT IS COMPOSED	
CC OF NINE SUBUNITS, SIX OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE	
CC A AND B CHAINS, AND THREE OF WHICH ARE DISULFIDE-LINKED DIMERS OF THE	
CC C CHAIN. IN ADDITION TO THE MAJOR A:B AND C:C DIMER BANDS,	
CC THE C CHAIN, IN ADDITION TO THE MAJOR A:B AND C:C DIMER BANDS,	
CC -1- SIMILARITY: CONTAINS 1 COLLAGENOUS DOMAIN.	
DR INTERPRO: IPR001073; C1q.	
DR PROSITE: PS01113; C1Q; PARTIAL.	
KW Complement pathway; Plasma; Hydroxylation; Glycoprotein; Collagen;	
KW Repeat.	
FT DISULFID 4 4	
FT FT	
FT MOD_RES 8 8	
FT MOD_RES 11 11	
FT MOD_RES 14 14	
FT MOD_RES 17 17	
FT NON_TER 28	
FT SEQUENCE 28 AA; 2712 MW; 643F01A24AD590EE CRC64;	
Query Match	17.7%; Score 34; DB 1; Length 28;
Best Local Similarity	50.0%; Pred. No. 1.6e+02;
Matches	5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY	25 NLGCVVSGLI 34
DB	1 NAGCVIGPM 10
RESULT 2	
AP65_CARMA	

```

ID AP65_CARMA STANDARD; PRT; 30 AA.
AC P82964;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antibacterial 6.5 kDa protein (Fragment).
OS Carcinus maenas (Common shore crab) (Green crab).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Crustacea;
OC Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata;
OC Brachyura; Eubrachyura; Portunoidae; Portunidae; Carcinus.
NCBI_TaxID=6759;
RN (1)
RP SEQUENCE, AND FUNCTION.
RC TISSUE=Hemocyte;
RX MEDLINE=97008941; PubMed=8856051;
RA Schnapp D., Kemp G.D., Smith V.J.;
RT "Purification and characterization of a proline-rich antibacterial
RT peptide, with sequence similarity to bactericin-7, from the haemocytes
RT of the shore crab, Carcinus maenas."
RL Eur. J. Biochem. 240:532-539 (1996).
CC -|- FUNCTION: STRONG ANTIMICROBIAL ACTIVITY AGAINST P. IMMOBILIS AND
CC -|- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED MW IS: 6.5 kDa.
CC -|- SIMILARITY: TO BOVINE BACTERICIN 7.
KW Antibiotic.
FT NON TER
SQ SEQUENCE 30 AA; 3305 MW; 6E2C2205934896C4 CRC64;

Query Match 16.7%; Score 32; DB 1; Length 30;
Best Local Similarity 45.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 PYGTLPGP 15
DB 8 PPRPRPGP 18

RESULT 3
HEMO_CHICK
ID HEMO_CHICK STANDARD; PRT; 34 AA.
AC P20057;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemopexin (Fragment).
GN HPX
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN (1)
RP SEQUENCE.
RX MEDLINE=88339942; PubMed=3421961;
RA Wellner D., Cheng K.C., Mueller-Eberhard U.;
RT "N-terminal amino acid sequences of the hemopexins from chicken, rat
RT and rabbit."
RL Biochem. Biophys. Res. Commun. 155:623-625 (1988).
CC -|- FUNCTION: BINDS HEME AND TRANSPORTS IT TO THE LIVER FOR BREAKDOWN
CC AND IRON RECOVERY, AFTER WHICH THE FREE HEMOPEXIN RETURNS TO THE
CC CIRCULATION.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- TISSUE SPECIFICITY: Synthesized by liver and secreted in plasma.
DR PIR: C31514; C31514.
DR InterPro: IPR000585; Hemopexin.
DR PROSITE: PS00024; HEMOPEXIN, PARTIAL.
KW Glycoprotein; Heme; Plasma; Repeat; Transport.
FT NON TER
SQ SEQUENCE 34 AA; 3598 MW; 07A2EAC27386C57 CRC64;

Query Match 16.1%; Score 31; DB 1; Length 34;
Best Local Similarity 55.6%; Pred. No. 5.2e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 4 HPYGLPVG 12
DB 15 HPHGAEPPG 23

RESULT 4
COOT_SARBU
ID COOT_SARBU STANDARD; PRT; 19 AA.
AC 009148;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NEB-colllostatin (Polliculostatin).
OS Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Oestroidea; Sarcophagidae; Sarcophaga.
NCBI_TaxID=7385;
RN (1)
RP SEQUENCE.
RX MEDLINE=95188911; PubMed=7883009;
RA Byllemans D., Froost P., Samijn B., Borovsky D., Grauwels L.,
RA Huybrechts R., van Damme J., van Beeumen J., de Loof A.;
RT "Neb-colllostatin, a second folliculostatin of the grey fleshfly,
RT Neobellieria bullata."
RL Eur. J. Biochem. 228:45-49 (1995).
CC -|- FUNCTION: HAS AN OSTOSTATIC ACTIVITY. IT INHIBITS YOLK SYNTHESIS SO
CC INHIBITING THE SUBSEQUENT YOLK DEPOSITION IN PREVITELLOGENIC
CC FOLLICLES.
CC -|- SIMILARITY: TO THE NONHELICAL REGIONS OF THE COLLAGEN FAMILY.
CC -|- CAUTION: NEB-COLLOSTATIN MAY BE PROTEOLYTICALLY CLEAVED FROM
CC COLLAGEN IV.
SQ SEQUENCE 19 AA; 1880 MW; 41B6AF5F0CB8251 CRC64;

Query Match 14.6%; Score 28; DB 1; Length 19;
Best Local Similarity 55.6%; Pred. No. 7.4e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 PYGTLPGP 13
DB 10 PIGPLVGP 18

RESULT 5
PA22_MICNI
ID PA22_MICNI STANDARD; PRT; 28 AA.
AC P21791;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase A2 isozyme 2 (EC 3.1.1.4) (Phosphatidylcholine
DE 2-acylhydrolase) (Fragment).
OS Micurus nigrocinctus (Central American coral snake) (Gargantilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Micurus.
NCBI_TaxID=8655;
RN (1)
RP SEQUENCE.
RC TISSUE=Venom;
RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Posseani L.D.;
RT "Isolation and characterization of three toxic phospholipases from
RT the venom of the coral snake Micurus nigrocinctus."
RL Toxicon 28:616-617 (1990).
CC -|- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
CC ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES. INHIBITS NEUROMUSCULAR
CC TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
CC TERMINI. ACT PRESYNAPTICALLY.
CC -|- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
CC acylglycerophosphocholine + a fatty acid anion.
CC -|- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -|- SUBCELLULAR LOCATION: Secreted.

```

CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.

DR PIR, B35948; B35948.
DR HSSB, P15445; 1A3D.
DR InterPro: IPR001211; PhospholipaseA2.
DR Pfam: PF00068; Phoslip; 1.
DR ProDom: PD000303; PhospholipaseA2; 1.
DR PROSITE: PS00118; PA2_HIS; PARTIAL.
DR PROSITE: PS00119; PA2_ASP; PARTIAL.
KW Hydrolase; Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
KW Multigene family.
FT NON TER 28
SQ SEQUENCE 28 AA; 3373 MW; 6979852DF2D718BC CRC64;

Query Match 14.6%; Score 28; DB 1; Length 28;
Best Local Similarity 55.6%; Pred. No. 1.1e+03;
Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 21 LRQNLGCT 29
| | | | |
DB 18 LFTNYGCT 26

RESULT 6

ID NU4M TRIRU STANDARD; PRT; 17 AA.
AC Q36834;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3) (Fragment).
GN ND4.
OS Trichophyton rubrum.
OC Mitochondrion.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;
OC Trichophyton.
NCBI_TaxID=5551;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IP.1817.89;
RA MEDLINE=96132111; PubMed=8593686;
RA de Bievre C., Dujon B.;
RT "Organisation of the mitochondrial genome of Trichophyton rubrum. DNA
RT sequence analysis of the ND4 gene, the ATPase subunit-6 gene, the
RT ribosomal RNA small-subunit gene, the ND6 gene, the COXIII gene, the
RT ATPase subunit-8 gene and six tRNA genes that correspond respectively
RT to the cytosine, lysine, glutamine, asparagine, isoleucine and
RT tryptophan isocodons.";
RT Curr. Genet. 28:553-559(1995).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----

DR EMBL; X88896; CAA61354.1; -
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON TER 1
SQ SEQUENCE 17 AA; 1834 MW; 225A167CE76F84BA CRC64;

Query Match 14.1%; Score 27; DB 1; Length 17;
Best Local Similarity 85.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 29 YVVGSLI 35
| | | | |
DB 9 YVVGSLI 15

RESULT 7

ID RL18 HALME STANDARD; PRT; 22 AA.
AC P50561;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18P (Hmel18) (Fragment).
GN RPL18P.
OS Halobacterium mediterranei (Haloflex mediterranei).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2252;
RX [1]
RP SEQUENCE.
RC STRAIN=DSM 1411;
RA MEDLINE=94229075; PubMed=8174557;
RA McDougall J., Wilmann-Liebold B.;
RT "Comparative analysis of the protein components from 5S rRNA, protein
RT complexes of halophilic archaeobacteria.";
RL Eur. J. Biochem. 221:779-785(1994).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
DR PIR, F33084; F33084.
KW RIBOSOMAL PROTEIN.
FT NON TER 22
SQ SEQUENCE 22 AA; 2773 MW; FE1F607FCAE9876D CRC64;

Query Match 14.1%; Score 27; DB 1; Length 22;
Best Local Similarity 44.4%; Pred. No. 1.2e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPEKTLR 22
| | | | |
DB 3 GPRYKVPWR 11

RESULT 8

ID RL18 HALVO STANDARD; PRT; 23 AA.
AC P50563;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18P (HVOL18) (Fragment).
GN RPL18P.
OS Halobacterium volcanii (Haloflex volcanii).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2246;
RX [1]
RP SEQUENCE.
RC STRAIN=DSM 3757;
RA MEDLINE=94229075; PubMed=8174557;
RA McDougall J., Wilmann-Liebold B.;
RT "Comparative analysis of the protein components from 5S rRNA, protein
RT complexes of halophilic archaeobacteria.";
RL Eur. J. Biochem. 221:779-785(1994).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
DR PIR, I33084; I33084.
FT NON TER 23
SQ SEQUENCE 23 AA; 2929 MW; 246E1F607FCAE987 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 23;
Best Local Similarity 44.4%; Pred. No. 1.3e+03;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPEKTLR 22
| | | | |
DB 3 GPRYKVPWR 11

RESULT 9

RL18_HALHA

ID RL18 HALHA STANDARD; PRT; 26 AA.
 AC PS0560;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 50S ribosomal protein L18p (HHA18) (Fragment).
 GN RPL18P.
 OS Halobacterium halobium.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 NC NCB1_TaxID=2242;
 RN (1)
 RP SEQUENCE.
 RC STRAIN=DSM 670;
 RX MEDLINE=94229075; PubMed=8174557;
 RA McDougall J., Wittmann-Liebold B.;
 RT "Comparative analysis of the protein components from 5S rRNA-protein
 complexes of halophilic archaeobacteria."
 RL Eur. J. Biochem. 221:779-785(1994).
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
 DR PIR: B33084; B33084.
 KW Ribosomal protein.
 FT NON_TER 26
 SQ SEQUENCE 26 AA; 3269 MW; D569CE746E1F607F CRC64;
 Query Match 14.1%; Score 27; DB 1; Length 26;
 Best Local Similarity 44.4%; Pred. No. 1.4e+03;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 GPEFKTLR 22
 |||:
 Db 3 GPRYKVMR 11

RESULT 10
 PA23_MICNI STANDARD; PRT; 27 AA.
 ID PA23 MICNI
 AC P21790;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phospholipase A2 isozyme 1 (EC 3.1.1.4) (Phosphatidylcholine
 2-acylhydrolase) (Fragment).
 OS Micrurus nigrocinctus (Central American coral snake) (Gargantilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Elapidae; Elapinae; Micrurus.
 NC NCB1_TaxID=8635;
 RN (1)
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Posanti L.D.;
 RT "Isolation and characterization of three toxic phospholipases from
 the venom of the coral snake Micrurus nigrocinctus."
 RL Toxicon 28:616-617(1990).
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
 ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS. INHIBITS NEUROMUSCULAR
 TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
 TERMINI. ACT PRESYNAPTICALLY.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 DR PIR: A35948; A35948.
 DR HSSP: P00598; 1POA.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF00068; Phoslip; 1.
 DR PROSITE: PS00118; PA2_HIS; PARTIAL.
 DR PROSITE: PS00119; PA2_ASP; PARTIAL.
 KW Hydrolyase, Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
 KW Multigene family.
 FT NON_TER 27

SQ SEQUENCE 27 AA; 3314 MW; 38637ECA600F49A0 CRC64;
 Query Match 14.1%; Score 27; DB 1; Length 27;
 Best Local Similarity 33.3%; Pred. No. 1.5e+03;
 Matches 4; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 18 KTLRFVQNLGCV 29
 |||:
 Db 15 RSVLEFMFGCV 26

RESULT 11
 PA23_MICNI STANDARD; PRT; 28 AA.
 ID PA23 MICNI
 AC P21792;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phospholipase A2 isozyme 3 (EC 3.1.1.4) (Phosphatidylcholine
 2-acylhydrolase) (Fragment).
 OS Micrurus nigrocinctus (Central American coral snake) (Gargantilla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Elapidae; Elapinae; Micrurus.
 NC NCB1_TaxID=8635;
 RN (1)
 RP SEQUENCE.
 RC TISSUE=Venom;
 RA Mochica-Morales J., Martin B.M., Zamudio F.Z., Posanti L.D.;
 RT "Isolation and characterization of three toxic phospholipases from
 the venom of the coral snake Micrurus nigrocinctus."
 RL Toxicon 28:616-617(1990).
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE 2-
 ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS. INHIBITS NEUROMUSCULAR
 TRANSMISSION BY BLOCKING ACETYLCHOLINE RELEASE FROM THE NERVE
 TERMINI. ACT PRESYNAPTICALLY.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 DR PIR: C35948; C35948.
 DR HSSP: P15445; 1A3D.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF00068; Phoslip; 1.
 DR PROSITE: PS00118; PA2_HIS; PARTIAL.
 DR PROSITE: PS00119; PA2_ASP; PARTIAL.
 KW Hydrolyase, Lipid degradation; Calcium; Presynaptic neurotoxin; Venom;
 KW Multigene family.
 FT NON_TER 28
 SQ SEQUENCE 28 AA; 3394 MW; 315FB012F69098B1 CRC64;
 Query Match 14.1%; Score 27; DB 1; Length 28;
 Best Local Similarity 80.0%; Pred. No. 1.5e+03;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 25 NLGCV 29
 |||:
 Db 22 NAGCV 26

RESULT 12
 PA2C_PSEPO STANDARD; PRT; 28 AA.
 ID PA2C PSEPO
 AC P20260;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phospholipase A2 (EC 3.1.1.4) (Pseudexin C chain) (Phosphatidylcholine
 2-acylhydrolase) (Fragment).
 OS Pseudexin porphyriacus (Red-bellied black snake).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Elapidae; Acantophinae; Pseudechis.
 OX NCBI_TaxID=8671;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Venom;
 RX MEDLINE=8938835; PubMed=2675391;
 RA Schmidt J.J., Middlebrook J.L.;
 RT "Purification, sequencing and characterization of pseudexin
 phospholipases A2 from Pseudechis porphyriacus (Australian
 red-bellied black snake).";
 RL Toxicon 27:805-818(1989).
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE
 CC 2-ACYL GROUPS IN 3-SN-PHOSPHOLIPIDS.
 CC -1- CATALYTIC ACTIVITY: Phosphatidylcholine + H(2)O = 1-
 CC acylglycerophosphocholine + a fatty acid anion.
 CC -1- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC PIR: C32416; C32416.
 DR HSSP: P00592; 2PH1.
 DR InterPro: IPR001211; PhospholipaseA2.
 DR Pfam: PF000068; Phoslip1.
 DR PRODOM: PD000303; PhospholipaseA2; 1.
 DR PROSITE: PS00118; PA2_HIS; PARTIAL.
 DR PROSITE: PS00119; PA2_ASP; PARTIAL.
 KW Hydrolase; Lipid degradation; Calcium; Multigene family; Venom.
 FT NON TER 28
 SQ SEQUENCE 28 AA; 3210 MW; 5089A7E85CAE0D5 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 28;
 Best Local Similarity 38.9%; Pred. No. 1.5e+03;
 Matches 7; Conservative 2; Mismatches 5; Indels 4; Gaps 1;
 Oy 12 GPGPEKTLRQNGCY 29
 Db 15 GSRPLFYT---DYGCY 28

RESULT 13
 RL18 HALCU
 ID RL18 HALCU STANDARD; PRT; 30 AA.
 AC P05970;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 50S ribosomal protein L18P (HCU18) (HL13) (Fragment).
 GN PPL18P.
 OS Halobacterium cutirubrum.
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxID=2242;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=79045279; PubMed=152199;
 RA Smith N., Matheson A.T., Yaguchi M., Willick G., Nazar R.N.;
 RT "The 5-S RNA-protein complex from an extreme halophile.
 RT Halobacterium cutirubrum. Purification and characterization.";
 RL Eur. J. Biochem. 89:501-509(1978).
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
 DR PIR: S07217; S07217.
 KW Ribosomal protein.
 FT NON TER 30
 SQ SEQUENCE 30 AA; 3624 MW; 3A50079B1569CB74 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 30;
 Best Local Similarity 44.4%; Pred. No. 1.7e+03;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 14 GPEKTLR 22
 Db 3 GPRYKVPWR 11

RESULT 14
 CY06 VIOOD
 ID CY06 VIOOD STANDARD; PRT; 31 AA.
 AC P58438;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cycloviolacin O6.
 OS Viola odorata (Sweet violet).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eustrodi 1; Malpighiales; Violaceae; Viola.
 OX NCBI_TaxID=97441;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=20069951; PubMed=10600388;
 RA Craik D.J., Daly N.L., Bond T., Waine C.;
 RT "Plant cyclotides: a unique family of cyclic and knotted proteins that
 RT defines the cyclic cysteine knot structural motif.";
 RL J. Mol. Biol. 294:1327-1336(1999).
 CC -1- FUNCTION: Probably participates in a plant defense mechanism.
 CC -1- PTM: This is a cyclic peptide.
 CC -1- CAUTION: This peptide is cyclic, its sequence was chosen to start
 CC at the position shown below by similarity to Oak1 (Kallata B1)
 CC whose DNA sequence is known.
 KW Multigene family.
 FT DISULFID 5 21
 FT DISULFID 9 23
 FT DISULFID 14 28
 SQ SEQUENCE 31 AA; 3208 MW; 30D6A52F72C6E68 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 31;
 Best Local Similarity 83.3%; Pred. No. 1.7e+03;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 7 GTLPVG 12
 Db 1 GTLPVG 6

RESULT 15
 CY0B VIOOD
 ID CY0B VIOOD STANDARD; PRT; 31 AA.
 AC P58433;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cycloviolacin O11.
 OS Viola odorata (Sweet violet).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eustrodi 1; Malpighiales; Violaceae; Viola.
 OX NCBI_TaxID=97441;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=20069951; PubMed=10600388;
 RA Craik D.J., Daly N.L., Bond T., Waine C.;
 RT "Plant cyclotides: a unique family of cyclic and knotted proteins that
 RT defines the cyclic cysteine knot structural motif.";
 RL J. Mol. Biol. 294:1327-1336(1999).
 CC -1- FUNCTION: Probably participates in a plant defense mechanism.
 CC -1- PTM: This is a cyclic peptide.
 CC -1- CAUTION: This peptide is cyclic, its sequence was chosen to start
 CC at the position shown below by similarity to Oak1 (Kallata B1)
 CC whose DNA sequence is known.
 KW Multigene family.
 FT DISULFID 5 21
 FT DISULFID 9 23
 FT DISULFID 14 28
 SQ SEQUENCE 31 AA; 3236 MW; 30D6A52F73EE568 CRC64;

Query Match 14.1%; Score 27; DB 1; Length 31;
 Oy 14 GPEKTLR 22
 Db 3 GPRYKVPWR 11

Best Local Similarity 83.3%; Pred. No. 1.7e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	7	GTLPVG	12
Db	1	GTLPVG	6

Search completed: July 16, 2003, 08:23:15
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:20:19 ; Search time 78 Seconds
(without alignments)
92.457 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Sequence: 1 YE VHPYGLPVGPPEFKTTLRVQNLGCVWSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 20244

```
Minimum DB seq length: 0
Maximum DB seq length: 35
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

1:  sp. archaea: *
2:  sp. bacteriia: *
3:  sp. fungi: *
4:  sp. inuman: *
5:  sp_invertebrate: *
6:  sp_mammal: *
7:  sp_minc: *
8:  sp_organelle: *
9:  sp_phages: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaea: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	46	24.0	34	15	091354	091354 human immun
2	46	24.0	34	15	091355	091355 human immun
3	46	24.0	34	15	091357	091357 human immun
4	46	24.0	34	15	078489	078489 human immun
5	45	23.4	34	15	078487	078487 human immun
6	45	23.4	35	15	091436	091436 human immun
7	45	23.4	35	15	091437	091437 human immun
8	45	23.4	35	15	090519	090519 human immun
9	45	23.4	35	15	076251	076251 human immun
10	44	22.9	35	15	091434	091434 human immun
11	44	22.9	35	15	091454	091454 human immun
12	44	22.9	35	15	080499	080499 human immun
13	44	22.9	35	15	010869	010869 human immun
14	44	22.9	35	15	076250	076250 human immun
15	44	22.9	35	15	076296	076296 human immun
16	43	22.4	34	15	091356	091356 human immun

17	43	22.4	34	15	Q70742	Q70742	human	immuru
18	43	22.4	34	15	Q70743	Q70743	human	immuru
19	43	22.4	34	15	Q70747	Q70747	human	immuru
20	43	22.4	35	15	O91349	O91349	human	immuru
21	43	22.4	35	15	O91350	O91350	human	immuru
22	43	22.4	35	15	O91351	O91351	human	immuru
23	43	22.4	35	15	O91430	O91430	human	immuru
24	43	22.4	35	15	O91433	O91433	human	immuru
25	43	22.4	35	15	O91435	O91435	human	immuru
26	43	22.4	35	15	O91438	O91438	human	immuru
27	43	22.4	35	15	O91439	O91439	human	immuru
28	43	22.4	35	15	O91440	O91440	human	immuru
29	43	22.4	35	15	O91446	O91446	human	immuru
30	43	22.4	35	15	O90503	O90503	human	immuru
31	43	22.4	35	15	O9WCW5	O9WCW5	human	immuru
32	43	22.4	35	15	O9OCU4	O9OCU4	human	immuru
33	43	22.4	35	15	O80465	O80465	human	immuru
34	43	22.4	35	15	O80466	O80466	human	immuru
35	43	22.4	35	15	O80467	O80467	human	immuru
36	43	22.4	35	15	Q78043	Q78043	human	immuru
37	43	22.4	35	15	Q78055	Q78055	human	immuru
38	43	22.4	35	15	Q78206	Q78206	human	immuru
39	43	22.4	35	15	Q78212	Q78212	human	immuru
40	43	22.4	35	15	Q78215	Q78215	human	immuru
41	43	22.4	35	15	O9WCW6	O9WCW6	human	immuru
42	43	22.4	35	15	Q70232	Q70232	human	immuru
43	43	22.4	35	15	Q79420	Q79420	human	immuru
44	43	22.4	35	15	Q76295	Q76295	human	immuru
45	43	22.4	35	15	Q77931	Q77931	human	immuru

ALIGNMENTS

RESULT 1					
ID	O91354	PRELIMINARY;	PRT;	34 AA.	
AC	O91354;				
DT	01-NOV-1998	(TREMBLrel . 08, Created)			
DR	01-NOV-1998	(TREMBLrel . 08, Last sequence update) .			
DT	01-JUN-2002	(TREMBLrel . 21, Last annotation update)			
DE	Envelope glycoprotein (Fragment) .				
CN	ENV.				
OS	Human immunodeficiency virus type 1.				
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.				
NCBI	_TaxId=11676;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=TK-22;				
RX	MEDLINE=96090117; PubMed=9430352;				
RA	Iida S., Gatanaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K., Kimura S., Iwamoto A., Oka S.;				
RT	"HIV type 1 v3 variation dynamics in vivo: long-term persistence of non-syncytium-inducing genotypes and transient presence of syncytium-inducing genotypes during the course of progressive AIDS."				
RL	AIDS Res. Hum. Retroviruses 13:1597-1609(1997).				
DR	EMBL; AB005326; BAA33244.1; -.				
KM	AIDS; Coat protein; Glycoprotein.				
FT	NON_TER	1			
FT	NON_TER	34			
SQ	SEQUENCE	34 AA; 3785 MW; A31972064FAD828F CRC64;			
Query Match		24.0%; Score 46; DB 15; Length 34;			
Best Local Similarity		42.1%; Pred. No. 15;			
Matches	8; Conservative	5; Mismatches	6; Indels	0; Gaps	0
OY	8 TLVPGPGPEFKTTLRVONTL 26				
	::: :::				
DB	11 SIPIGPGRAFYTNIIGNI 29				
RESULT 2					
O91355					

```

ID 091355 PRELIMINARY; PRT; 34 AA.
AC 091355;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-22;
RX MEDLINE=98090117; PubMed=9430252;
RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,
RA Kimura S., Iwamoto A., Oka S.;
RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of
RT non-synctium-inducing genotypes and transient presence of cynctium-
RT inducing genotypes during the course of progressive AIDS.";
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
DR EMBL; AB05327; BAA33245.1; -
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3785 MW; A31972064FAD828F CRC64;

Query Match 24.0%; Score 46; DB 15; Length 34;
Best Local Similarity 42.1%; Pred. No. 15;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 8 TLVPGPPEFKTTLRVONL 26
DB 11 SIPIGGRAFTYTTNIGNI 29

```

```

RESULT 3
ID 091357 PRELIMINARY; PRT; 34 AA.
AC 091357;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-22;
RX MEDLINE=98090117; PubMed=9430252;
RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,
RA Kimura S., Iwamoto A., Oka S.;
RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of
RT non-synctium-inducing genotypes and transient presence of cynctium-
RT inducing genotypes during the course of progressive AIDS.";
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
DR EMBL; AB05329; BAA33247.1; -
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3811 MW; A303C2064FAD828F CRC64;

```

```

Query Match 24.0%; Score 46; DB 15; Length 34;
Best Local Similarity 42.1%; Pred. No. 15;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 8 TLVPGPPEFKTTLRVONL 26
DB 11 SIPIGGRAFTYTTNIGNI 29

```

RESULT 4

```

Q78489
ID Q78489 PRELIMINARY; PRT; 34 AA.
AC Q78489;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Viral sample FLO5R3C (Florida local control 01), partial env cds, V3
DE region (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC Zhang L.O., Leigh-Brown A.J.;
RT Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271245; PubMed=1589796;
RA Ou C.-Y., Ciesielski C.A., Myers G., Bantee C.I., Luo C.C.,
RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,
RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., Curran J.W.,
RA Jaffe H.W.;
RT "Molecular Epidemiology of HIV Transmission in a Dental Practice.";
RL Science 256:1165-1171(1992).
DR EMBL; M92128; AAA44576.1; -
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3786 MW; A93F03A44FAD828F CRC64;

```

```

Query Match 24.0%; Score 46; DB 15; Length 34;
Best Local Similarity 42.1%; Pred. No. 15;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 8 TLVPGPPEFKTTLRVONL 26
DB 11 SIPIGGRAFTYTTNIGNI 29

```

```

RESULT 5
ID Q78487 PRELIMINARY; PRT; 34 AA.
AC Q78487;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Viral sample FLO5R3A (Florida local control 01), partial env cds, V3
DE region (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC Zhang L.O., Leigh-Brown A.J.;
RT Submitted (Apr-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271245; PubMed=1589796;
RA Ou C.-Y., Ciesielski C.A., Myers G., Bantee C.I., Luo C.C.,
RA Korber B.T.M., Mullins J.I., Schochetman G., Berkelman R.L.,
RA Economou A.N., Witte J.J., Furman L.J., Satten G.A., Curran J.W.,
RA Jaffe H.W.;
RT "Molecular Epidemiology of HIV Transmission in a Dental Practice.";
RL Science 256:1165-1171(1992).
DR EMBL; M92127; AAA44575.1; -
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 34 34
SQ SEQUENCE 34 AA; 3818 MW; A93F02E148C0428F CRC64;

```

Query Match 23.4%; Score 45; DB 15; Length 34;

Best Local Similarity 42.1%; Pred. No. 21;
Matches 8; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
QY 8 TLPVGPGEPEKTTLRVONT 26
:::|||||:::
DB 11 SIPMGPGRAFYTTTROI 29

RESULT 6

091436 PRELIMINARY; PRT; 35 AA.
AC 091436;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-2;
RA MEDLINE=96090117; PubMed=9430252;
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,
RA Kimura S., Iwamoto A., Oka S.;
RT "HIV type 1 V3 variation dynamics in vivo:long-term persistence of
RT non-synctium-inducing genotypes and transient presence of cynctium-
RL inducing genotypes during the course of progressive AIDS.";
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
DR EMBL; AB005421; BAA3338.1; -;
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT 35
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;
Best Local Similarity 43.8%; Pred. No. 22;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLPVGPGEPEKTTLRV 23
:::|||||:::
DB 11 SIPMGPGRAFYTTROI 26

RESULT 7

091437 PRELIMINARY; PRT; 35 AA.
AC 091437;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-2;
RA MEDLINE=96090117; PubMed=9430252;
RA Ida S., Gatnaga H., Shioda T., Nagai Y., Kobayashi N., Shimada K.,
RA Kimura S., Iwamoto A., Oka S.;
RT "HIV type 1 V3 variation dynamics in vivo:long-term persistence of
RT non-synctium-inducing genotypes and transient presence of cynctium-
RL inducing genotypes during the course of progressive AIDS.";
RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
DR EMBL; AB005422; BAA3339.1; -;
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT 35
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;
Best Local Similarity 43.8%; Pred. No. 22;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 8 TLPVGPGEPEKTTLRV 23
:::|||||:::
DB 11 SIPMGPGRAFYTTROI 26

RESULT 8

090519 PRELIMINARY; PRT; 35 AA.
AC 090519;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=745;
RA Vallejo A., Mas A., Heredia A., Altisent C., Lorenzo I., Soriano V.,
RA Hewlett I.K.;
RT "Virologic characteristics and genetic analysis of the V3 loop and the
RT nef gene of an HIV-1-infected hemophilic cohort with long-term non-
RT progressive infection.";
RL submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF082373; AAC34519.1; -;
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT 35
SQ SEQUENCE 35 AA; 3878 MW; 9F5A59F6B78C81BE CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 8 TLPVGPGEPEKTTLRV 23
:::|||||:::
DB 11 SIPMGPGRAFYTTROI 26

RESULT 9

076251 PRELIMINARY; PRT; 35 AA.
AC 076251;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE V3 loop (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Schreiber M.G.;
RT "Loss of V3 domain specific antibodies to selected HIV-1 variants over
RT time is associated with viral escape and selection in vivo.";
RL submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z50845; CAA90698.1; -;
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT 35
SQ SEQUENCE 35 AA; 3956 MW; 9F5A5B3A688BF77E CRC64;

Query Match 23.4%; Score 45; DB 15; Length 35;
Best Local Similarity 43.8%; Pred. No. 22;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 8 TLVGGPPEFKTLRV 23
 Db 11 SIPIGGRAFYTTROI 26

RESULT 10

091434 PRELIMINARY; PRT; 35 AA.
 AC 091434;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_Taxid=11676;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=TK-2;
 RX MEDLINE=98090117; PubMed=9430252;
 RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K., Kimura S., Iwamoto A., Oka S.,
 RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of non-synctium-inducing genotypes and transient presence of cynctium-inducing genotypes during the course of progressive AIDS.";
 RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
 DR EMBL; AB005419; BAA3336.1; -
 KM AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT 35
 SQ SEQUENCE 35 AA; 3831 MW; 9F5ASB2A698BE26B CRC64;
 OY 8 TLVGGPPEFKTT 20
 Db 11 SIPIGGRAFYTT 23

RESULT 11

091454 PRELIMINARY; PRT; 35 AA.
 AC 091454;
 DT 01-NOV-1998 (TRENBLrel. 08, Created)
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_Taxid=11676;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=TK-2;
 RX MEDLINE=98090117; PubMed=9430252;
 RA Ida S., Gatanga H., Shioda T., Nagai Y., Kobayashi N., Shimada K., Kimura S., Iwamoto A., Oka S.,
 RT "HIV type 1 V3 variation dynamics in vivo: long-term persistence of non-synctium-inducing genotypes and transient presence of cynctium-inducing genotypes during the course of progressive AIDS.";
 RL AIDS Res. Hum. Retroviruses 13:1597-1609(1997).
 DR EMBL; AB005440; BAA3336.1; -
 KM AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT 35
 SQ SEQUENCE 35 AA; 4026 MW; 9E4A4A3A68BF77E CRC64;

Query Match 22.9%; Score 44; DB 15; Length 35;
 Best Local Similarity 46.7%; Pred. No. 30;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 9 LPVGGPPEFKTLRV 23
 Db 12 IPIGGRAFYTTROI 26

RESULT 12

080499 PRELIMINARY; PRT; 35 AA.
 AC 080499;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Envelope protein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_Taxid=11676;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90364416; PubMed=2392685;
 RA Larosa G.J., Davide J.P., Weinhold K., Waterbury J.A., Profy A.T., Lewis J.A., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Holley L.H., Karplus M., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant";
 RL Science 249:932-935(1990).
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9118022; PubMed=1990444;
 RA Larosa G.J., Davide J.P., Weinhold K., Waterbury J.A., Profy A.T., Lewis J.A., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Holley L.H., Karplus M., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: Corrections and clarifications.";
 RL Science 251:811-811(1991).
 RN 13
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91361090; PubMed=1887238;
 RA Larosa G.J., Weinhold K., Profy A.T., Langlois A.J., Dreesman G.R., Boswell R.N., Shadduck P., Bolognesi D.P., Matthews T.J., Emini E.A., Putney S.D.;
 RT "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: Further clarifications.";
 RL Science 253:1146-1146(1991).
 DR EMBL; M61532; AAA45222.1; -
 KM AIDS; Coat protein; Glycoprotein.
 FT NON_TER 1
 FT 35
 SQ SEQUENCE 35 AA; 4022 MW; 0780966F52E750D0 CRC64;

Query Match 22.9%; Score 44; DB 15; Length 35;
 Best Local Similarity 34.6%; Pred. No. 30;
 Matches 9; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

OY 3 VHPYGT---LPVGGPPEFKTLRVQ 24
 Db 2 IRPHNTIRRHIGGRAFYTTROI 27

RESULT 13

010869 PRELIMINARY; PRT; 35 AA.
 AC 010869;
 DT 01-JUL-1997 (TRENBLrel. 04, Created)
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Envelope glycoprotein (Fragment).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OC NCBI_Taxid=11676;

RM [1]
RC SEQUENCE FROM N.A.
RN STRAIN=LIFI;
RX MEDLINE=97186421; PubMed=9115816;
RA Mang B., Ge Y.C., Palasanthiran P., Ziegler J., Bolton W., Xiang S.H.,
RA Dwyer D.E., Cunningham A.L., Sakena N.K.,
RT "HIV type 1 V3 loop sequences derived from peripheral blood of
transmitting mothers, their infants, and nontransmitting mothers
differ in their crown octapeptide motifs."
RL AIDS Res. Hum. Retroviruses 13:275-279 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LIFI;
RA Blouin J.C.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U66657; AAB58149.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3718 MW; 9F53DA6E98B8B8C CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 36.0%; Pred. No. 30;
Matches 9; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

Oy 8 TLPGPGPEFKTTLR---VQNLGC 28
Db 11 SIPIGPGRAFTTGTGIIIGDIRQAGC 35

RESULT 14
O76250 PRELIMINARY; PRT; 35 AA.
AC Q76250;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE V3 loop (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Schreiber M.G.;
RT "Loss of V3 domain specific antibodies to selected HIV-1 variants over
time is associated with viral escape and selection in vivo."
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z50844; CAA90697.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3898 MW; 8A1B0F25558BF77E CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 46.7%; Pred. No. 30;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 9 LPVGPPEFKTLRV 23
Db 12 IPIGPGRAFTTQTQI 26

RESULT 15
O76296 PRELIMINARY; PRT; 35 AA.
AC Q76296;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE V3 loop (Fragment).
GN ENV.
OS Human immunodeficiency virus type 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94246736; PubMed=8189527;
RA Schreiber M.G., Petersen H., Wachsmuth C., Muller H., Hufert F.T.,
RA Schmitz H.;
RT "Antibodies of symptomatic human immunodeficiency virus type 1-
infected individuals are directed to the V3 domain of noninfectious
RT and not of infectious virions present in autologous serum."
RL J. Virol. 68:3908-3916 (1994).
DR EMBL; Z15163; CAA78862.1; -.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 1
FT SEQUENCE 35 AA; 3775 MW; 9F5B905B698BF77E CRC64;
SQ

Query Match 22.9%; Score 44; DB 15; Length 35;
Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 8 TLPGPGPEFKTT 20
Db 11 SIPIGPGSAFTT 23

Search completed: July 16, 2003, 08:24:40
Job time : 79 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 08:14:48 / Search time 68 seconds
(without alignments)
68.585 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192
Sequence: 1 YEVHPYGTLPVGPGEFKTLRVQNLGCVVSGHI 35

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 375850

Minimum DB seq length: 0
Maximum DB seq length: 35

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	24.0	33	16	AA68713
2	46	24.0	33	18	AAW25883
3	46	24.0	33	19	AAW67399
4	46	24.0	33	20	AAV39728
5	46	24.0	33	20	AAW98900
6	46	24.0	33	16	AA677763
7	45.5	23.7	33	16	AA68712
8	45.5	23.7	33	18	AAW25882
9	45.5	23.7	33	19	AAW67398
10	45.5	23.7	33	20	AAV39727

11	45.5	23.7	33	20	AAW98899
12	45.5	23.7	35	16	AA677762
13	44	22.9	24	11	AA604467
14	44	22.9	35	21	AA610705
15	43.5	22.7	33	16	AA68710
16	43.5	22.7	33	18	AAW25880
17	43.5	22.7	33	19	AAW67396
18	43.5	22.7	33	20	AAV39725
19	43.5	22.7	33	20	AAW98897
20	43.5	22.7	35	16	AA677760
21	43	22.4	18	16	AAV04445
22	43	22.4	18	16	AA68679
23	43	22.4	18	18	AAW25849
24	43	22.4	18	19	AAW67365
25	43	22.4	18	20	AAV39698
26	43	22.4	18	20	AAW99973
27	43	22.4	24	14	AA68170
28	43	22.4	33	16	AA68670
29	43	22.4	33	16	AA68709
30	43	22.4	33	18	AAW25879
31	43	22.4	33	18	AAW25840
32	43	22.4	33	19	AAW67395
33	43	22.4	33	19	AAW67356
34	43	22.4	33	20	AAV39724
35	43	22.4	33	20	AAV39762
36	43	22.4	33	20	AAW98896
37	43	22.4	33	20	AAW99964
38	43	22.4	34	16	AA68674
39	43	22.4	34	18	AAW25844
40	43	22.4	34	19	AAW67360
41	43	22.4	34	20	AAV39765
42	43	22.4	34	20	AAW99968
43	43	22.4	35	13	AA621089
44	43	22.4	35	13	AA621090
45	43	22.4	35	13	AA621134

ALIGNMENTS

RESULT 1	AA68713
ID	AA68713 standard; peptide; 33 AA.
XX	AA68713;
AC	AA68713;
XX	07-SEP-1995 (first entry)
XX	PRI/T5 chimera, CLTB-159.
DE	
XX	T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;
KW	gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.
XX	
OS	Human immunodeficiency virus type 1.
XX	
PH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..18
FT	/label= PRI
FT	/note= "B cell epitope"
FT	19..33
FT	/label= T5
FT	/note= "T cell epitope"
XX	
PN	WO9429339-A.
XX	
PD	22-DEC-1994.
XX	
PF	08-JUN-1994; 94WO-CA00317.
XX	
PR	09-JUN-1993; 93US-0073378.
XX	
PA	(CONN-) CONNAUGHT LAB LTD.
XX	

HIV-1 vaccine synt
T5/PRI chimera lys
Human immunodefici
HIV-1 isolate PI-9
PRI/p24H chimera,
Chimeric T/B cell
HIV-1 peptide epit
HIV1 chimeric pep
HIV-1 vaccine synt
PRI/p24H chimera 1
Peptide CLTB-PRI.
Consensus B cell e
HIV-1 env protein
HIV-1 V3 loop pep
HIV1 chimeric pep
HIV-1 vaccine synt
V3 loop peptide D2
p24E/V3 consensus
Chimeric T/B cell
Chimeric T/B cell
HIV-1 peptide epit
HIV-1 peptide epit
HIV1 chimeric pep
HIV-1 vaccine synt
HIV-1 vaccine synt
T1/V3 consensus se
Chimeric T/B cell
HIV-1 peptide epit
HIV1 chimeric pep
HIV-1 vaccine synt
PND E815-2. Human
PND E815-3. Human
PND E828-1. Huma

Db 5 SIPGPGAFYTTGYKKYKVKIEPLG 30

AC AAW25883;

DE Chimeric T/B cell epitope peptide CLTB-159.

PN US5639854-A.

PF 09-JUN-1993; 93US-0073378.

PA (CONN-) CONNAUGHT LAB LTD.
VY

DR WPI; 1997-332082/30.
XX

PS Example 1; Column 23-24; 41pp; English.

CC The invention relates to new synthetic peptides comprising at least one
CC amino acid sequence comprising an HIV gag protein T-cell epitope linked
CC at its C- or N-terminus to an amino acid sequence comprising a B-cell
CC epitope of the V3 loop of an HIV env protein, which can be used to
CC generate vaccines against HIV-1. The T-cell epitope sequence is pref.

Db 5 SIPIGPGRAFYTTGYKKKKVVKIEPLG 30

DT 25-JAN-1999 (first entry)

OS Synthetic.

OS Human immunodeficiency virus type 1.

PF 05-JUN-1995; 95US-0464329.

PR 05-JUN-1995; 95US-0464329.

PI : Chong P, Klein MH, Sia CDY;

DR WPI; 1998-556461/47

PS Disclosure; Column 23-24; 40pp; English

CC The invention relates to a novel immunogenic composition for use in
CC vaccines for the treatment of HIV-1 comprising an HIV-1-derived T-cell
CC epitope linked to an HIV-1-derived B-cell epitope. The T-cell epitopes
CC are generally designed based on the p24 core protein and the B-cell
CC epitopes from the V3 loop of the gp120 protein from various HIV-1
CC strains. This peptide corresponds to a fusion of the T5 epitope
CC (AAW6739) and the V3 loop B-cell epitope (AAW67365) which is a
CC consensus V3 loop peptide of the New York and Amsterdam HIV-1 isolates.

Sequence 33 AA;

Query Match	24.0%;	Score 46;	DB 19;	Length 33;
Best Local Similarity	34.6%;	Pred. No. 30;		

Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

Qy 8 TLPVGPGEFKT-----LRVQNLG 27
 ::|||::|||
 Db 5 SIPIGGRAFYTTGYKVKVKEIPLG 30

RESULT 4
 AAY39728
 ID AAY39728 standard; peptide; 33 AA.
 XX AAY39728;
 AC
 XX 26-NOV-1999 (first entry)
 DT
 XX HIV1 chimeric peptide CLTB-159.
 DE
 XX HIV, vaccine; immunogenic composition; T cell epitope; B cell epitope;
 KM infection; antibody; antiviral.
 XX
 XX Human immunodeficiency virus type 1.
 OS
 XX US5951986-A.
 PN
 XX 14-SEP-1999.
 PD
 XX 06-JUN-1995; 95US-0467881.
 PF
 XX 09-JUN-1994; 94US-0257528.
 PR
 XX 09-JUN-1993; 93US-0073378.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 PI Klein MH, Chong P, Sia CDY;
 XX WPI; 1999-550482/46.
 DR
 XX Immunogenic composition containing synthetic fusion polypeptides
 PT containing both the T and B cell epitopes of the human immunodeficiency
 PT virus, useful antigens in producing vaccines -
 XX
 XX Example 1; Column 25-26; 43pp; English.
 PS
 XX This sequence represents a fragment of a HIV1 protein, and can be used in
 CC the immunogenic composition of the invention. The composition comprises a
 CC synthetic fusion polypeptide which includes a sequence encoding 1 or more
 CC T cell epitopes and a sequence encoding 1 or more B cell epitopes and a
 CC carrier. Both the T cell and B cell epitopes are derived from HIV
 CC proteins. The compositions are useful as vaccines against HIV infection.
 CC The composition induces HIV-1-specific polyclonal antibodies that are
 CC opsonising and antiviral. The peptide components may be selected to
 CC induce a response against different viral isolates and in subjects who
 CC recognise different T cell epitopes.
 CC
 XX Sequence 33 AA;
 SQ

Query Match 24.0%; Score 46; DB 20; Length 33;
 Best Local Similarity 34.6%; Pred. No. 30;
 Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

Qy 8 TLPVGPGEFKT-----LRVQNLG 27
 ::|||::|||
 Db 5 SIPIGGRAFYTTGYKVKVKEIPLG 30

RESULT 5
 AAW98900
 ID AAW98900 standard; peptide; 33 AA.
 XX AAW98900;
 AC
 XX 05-MAY-1999 (first entry)
 DT
 XX

DE HIV-1 vaccine synthetic peptide SEQ ID NO:84.
 XX HIV-1; human immunodeficiency virus; vaccine; T-cell epitope;
 KM gag protein; B-cell epitope; gp41 protein; chimeric; infection.
 XX
 XX Synthetic.
 OS Human immunodeficiency virus type 1.
 XX
 XX US5876731-A.
 PN
 XX 02-MAR-1999.
 PD
 XX 05-JUN-1995; 95US-0462507.
 PF
 XX 09-JUN-1994; 94US-0257528.
 PR
 XX 09-JUN-1993; 93US-0073378.
 PR
 XX 05-JUN-1995; 95US-0462507.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 PA
 PI Chong P, Klein MH, Sia CDY;
 XX WPI; 1999-189590/16.
 DR
 XX Synthetic chimeric HIV polypeptides - comprising gag protein T-cell
 PT epitope linked to gp41 B-cell epitope
 PT
 XX Example 1; Column 63-64; 41pp; English.
 PS
 XX The present invention describes a synthetic peptide comprising an amino
 CC acid sequence containing a T-cell epitope of an HIV gag protein linked
 CC at its C terminus to an amino acid sequence containing a B-cell epitope
 CC of an HIV gp41 protein and containing the amino acid sequence:
 CC X1LKDWX2; where X1 = E, A, G or Q, and X2 = A or T, or an amino acid
 CC sequence capable of eliciting an HIV-specific antiserum and recognizing
 CC the sequence X1LKDWX2. The synthetic peptide is useful in vaccines
 CC against HIV infection and in diagnostic applications. AAW98892 to
 CC AAW98906, and AAW98999 to AAW99899 represent synthetic peptides from the
 CC present invention.
 CC
 XX Sequence 33 AA;
 SQ

Query Match 24.0%; Score 46; DB 20; Length 33;
 Best Local Similarity 34.6%; Pred. No. 30;
 Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

Qy 8 TLPVGPGEFKT-----LRVQNLG 27
 ::|||::|||
 Db 5 SIPIGGRAFYTTGYKVKVKEIPLG 30

RESULT 6
 AAR77763
 ID AAR77763 standard; peptide; 35 AA.
 XX AAR77763;
 AC
 XX 08-SEP-1995 (first entry)
 DT
 XX PRI/T5 chimera lysine branched peptide.
 DE
 XX T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;
 KM gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.
 XX
 XX Human immunodeficiency virus type 1.
 OS
 XX Key
 XX Location/Qualifiers
 FH 1..18
 FT Peptide
 FT /label= PRI
 FT /note= "B cell epitope"
 FT 19..33
 FT /label= T5
 FT /note= "T cell epitope"

```
FT Modified-site 34 /note= "Modified at the epsilon N by the sequence given
FT /note= "in AAR68713, which represents the same sequence
FT as bases 1-33 of this sequence"
FT Modified-site 35 /note= "Modified at the epsilon N with a Lys residue
FT which is substituted on the alpha and epsilon
FT N with the sequence given in AAR68713"
PN WO9429339-A.
XX
XX 22-DEC-1994.
PD
XX 08-JUN-1994; 94WO-CA00317.
PF
XX 09-JUN-1993; 93US-0073378.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Chong P, Klein MH, Sia CDY;
PI
XX WPI; 1995-036400/05.
DR
XX
XX Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell
PT epitope of gag protein linked to B-cell epitope of V3 loop
PT protein of an HIV-1 isolate
XX
XX Claim 4; Page 46; 69pp; English.
XX
XX This sequence represents a multimeric lysine branched peptide which
CC contains four copies a peptide monomer comprising the T-cell epitope
CC derived from the HIV-1 core protein, T5, linked to the B-cell epitope,
CC PRI which represents a consensus sequence derived from New York and
CC Amsterdam HIV-1 isolates. The entire multimer contains four copies of
CC the same sequence linked through Lys residues. The B-cell epitope may
CC be derived from HIV-1 proteins esp. gp160, gag and pol proteins. These
CC chimeric peptides may then be used in the production of HIV-1 vaccines.
CC The linear and multimeric peptides may be used for the treatment of AIDS
CC by acting to displace the binding of HIV virus to human or animal cells
CC or by disturbing the 3D organisation of the virus.
XX
SQ Sequence 35 AA;
Query Match 24.0%; Score 46; DB 16; Length 35;
Best Local Similarity 34.6%; Pred. No. 32;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;
QY 8 TLPGGPEPKTT-----LRVQNIIG 27
Db 5 SIPIGPGRAFYTTGYKVKVKTIEPLG 30
RESULT 7
AAR68712 ID AAR68712 standard; peptide; 33 AA.
XX
XX AAR68712;
AC
XX
XX 07-SEP-1995 (first entry)
DT
XX
XX T5/PRI chimera, CLTB-158.
DE
XX
XX T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;
KW gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.
XX
XX Human immunodeficiency virus type 1.
OS
XX
XX Key Location/Qualifiers
FH 1..15 /label= T5
FT /note= "T cell epitope"
FT 17..33 /label= PRI
FT Peptide
```

```
FT /note= "B cell epitope"
XX
XX WO9429339-A.
PN
XX 22-DEC-1994.
PD
XX 08-JUN-1994; 94WO-CA00317.
PF
XX 09-JUN-1993; 93US-0073378.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Chong P, Klein MH, Sia CDY;
PI
XX WPI; 1995-036400/05.
DR
XX
XX Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell
PT epitope of gag protein linked to B-cell epitope of V3 loop
PT protein of an HIV-1 isolate
XX
XX Claim 4; Page 46; 69pp; English.
XX
XX This sequence represents the T-cell epitope derived from the HIV-1
CC core protein, T5, linked to consensus B-cell epitope, PRI which
CC represents a consensus sequence derived from New York and Amsterdam
CC HIV-1 isolates. Chimeric peptides such as this, may be used in the
CC production of HIV-1 vaccines. These peptide sequences may also be
CC used in the production of multimeric peptides in which the peptides
CC are C-terminally modified by the addition of a Lys residue which is
CC modified on its epsilon amino acid to carry an additional copy of
CC the peptide molecule. The linear and multimeric peptides may be
CC used for the treatment of AIDS by acting to displace the binding of
CC HIV virus to human or animal cells or by disturbing the 3D
CC organisation of the virus.
XX
SQ Sequence 33 AA;
Query Match 23.7%; Score 45.5; DB 16; Length 33;
Best Local Similarity 34.6%; Pred. No. 36;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;
QY 2 EVHPYG-----TLPGGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGPGRAFYTT 32
RESULT 8
AAW25882 ID AAW25882 standard; peptide; 33 AA.
XX
XX AAW25882;
AC
XX
XX 22-OCT-1997 (first entry)
DT
XX
XX Chimeric T/B cell epitope peptide CLTB-158.
DE
XX
XX HIV; human immunodeficiency virus; gag; T-cell; B-cell; epitope; env;
KW V3 loop; vaccine; determinant; chimeric.
XX
XX Synthetic.
OS
XX
XX US5639854-A.
PN
XX 17-JUN-1997.
PD
XX 09-JUN-1993; 93US-0073378.
PF
XX 09-JUN-1994; 94US-0257528.
PR
XX 09-JUN-1993; 93US-0073378.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX Chong P, Klein MH, Sia CDY;
PI
```


DE Human immunodeficiency virus peptide RP150.
XX
XX HIV isolate AVR-2; peptide RP150; principal neutralising domain;
KW antibodies; diagnosis; prophylaxis; therapy; AIDS.
XX
XX Synthetic.
OS
XX WO9003984-A.
PN
XX
XX 19-APR-1990.
PD
XX
XX 29-SEP-1989; 89WO-US04302.
PF
XX
XX 19-SEP-1989; 89US-0407663, US-252949; WO-U04302.
PR
XX
XX (REPL-) REPLIGEN CORP.
PA
XX
XX Rusche JR, Putney SD, Javaherian K, Farley J, Grimalia R, Lynn D,
PI Petro-Breyer J;
XX
XX WPI; 1990-147824/19.
DR
XX
XX Principal neutralising domain of HIV variants - used for producing
PT peptide(s) and antibodies for diagnosis; prophylaxis; and/or therapy
PT of HIV infection.
XX
XX
XX Claim 8 (81); Page 77; 108pp; English.
PS
XX
XX Peptide RP150 comprises segments of the Principal Neutralising Domain
CC (Envelope protein) from isolate AVR-2. The last Cys residue is
CC added for the purpose of crosslinking to carrier proteins. Cysteine
CC residues may be added, so that the residues at or near both ends form a
CC disulfide bond, giving peptide a loop-like configuration, which can be
CC utilized to enhance the immunogenic properties of the peptides. Protein
CC is capable of eliciting, and/or binding with, neutralising Abs. The
CC neutralising domain is bounded by cysteine residues which occur at
CC positions 296 and 331. Peptides can be used as immunogens or screening
CC reagents to generate or identify poly- or monoclonal antibodies.
CC See also AAR0427-R04506 and AAQ04273-Q04279.
CC
XX
XX Sequence 24 AA;
SQ
Query Match 22.9%; Score 44; DB 11; Length 24;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 9; Conservative 3; Mismatches 4; Indels 2; Gaps 1;
OY 11 VGPGEFKTLRVQNLGC 28
Db 9 IGPGRFHTTGRi-IGC 24
RESULT 14
AAB10706
ID AAB10706 standard; Protein; 35 AA.
AC
XX
XX AAB10706;
AC
XX
XX 26-JAN-2001 (first entry)
DT
XX
XX HIV-1 isolate PI-932 gp120 protein V3 loop region fragment.
DB
XX
XX Viral protein; vaccine; anti-viral; anti-HIV; therapy; infection;
KW gp120; V3 loop.
XX
XX
XX Human immunodeficiency virus type 1.
OS
XX
XX WO200047223-A2.
PN
XX
XX 17-AUG-2000.
PD
XX
XX 03-DEC-1999; 99WO-EP09759.
PF
XX
XX 12-FEB-1999; 99DE-1007485.
PR

XX
XX (STRA-) STRATHMANN & CO AG.
PA
XX
XX Schreiber M;
PI
XX
XX WPI; 2000-549084/50.
DR
XX
XX
XX Viral vaccine comprises a mixture of protein sequence variants of a
PT single viral protein, which is useful for prevention and therapy of
PT viral infections, especially HIV, in humans
XX
XX
XX Disclosure; Fig 1; 79pp; German.
ES
XX
XX This invention describes a novel protein vaccine which comprises a
CC mixture of viral proteins, characterized in that the molecules are
CC sequence variants of a single viral (partial) protein. The products of
CC the invention have anti-viral and anti-HIV activity. Mixtures of
CC structurally different viral proteins, that are sequence variants of a
CC single protein are useful for production of vaccines for the prevention
CC and/or therapy of viral infections in humans. The vaccines are especially
CC useful for prevention and/or therapy of human immunodeficiency virus
CC (HIV) infection in humans. The vectors and host cells are useful for the
CC expression of the protein/DNA mixtures, which are also useful for the
CC preventing and/or therapy of viral infection. This sequence represents
CC a gp120 protein V3 loop fragment from HIV-1 isolate PI-932.
CC
XX
XX Sequence 35 AA;
SQ
Query Match 22.9%; Score 44; DB 21; Length 35;
Best Local Similarity 30.8%; Pred. No. 61;
Matches 8; Conservative 7; Mismatches 7; Indels 4; Gaps 1;
OY 3 VHPYGT---LPVGPGEFKTLRVQ 24
Db 2 IRPHNTVDRHIGPGRSFHTTKIK 27
RESULT 15
AAR68710
ID AAR68710 standard; peptide; 33 AA..
AC
XX
XX AAR68710;
AC
XX
XX 07-SEP-1995 (first entry)
DT
XX
XX PRI/P24H chimera, CTRB-157.
DE
XX
XX T-cell; epitope; HIV-1; core protein; p24E; B-cell; antigen;
KW gp160; gag; pol; vaccine; multimeric peptide; AIDS; 3D organisation.
KW
XX
XX Human immunodeficiency virus type 1.
OS
XX
XX
XX Key Location/Qualifiers
FH 1..18
FT /label= PRI
FT /note= "B cell epitope"
FT 19..33
FT /label= P24H
FT /note= "T cell epitope"
XX
XX
XX WO9429339-A.
PN
XX
XX 22-DEC-1994.
PD
XX
XX 08-JUN-1994; 94WO-CA00317.
PF
XX
XX 09-JUN-1993; 93US-0073378.
PR
XX
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX
XX Chong P, Klein MH, Sia CDY,
PI
XX
XX WPI; 1995-036400/05.
DR

XX Novel tandem synthetic HIV-1 peptide(s) - comprising T-cell
 PT epitope of gag protein linked to B-cell epitope of V3 loop
 PT protein of an HIV-1 isolate

XX Claim 4; Page 45; 69pp; English.

XX This sequence represents the T-cell epitope derived from the HIV-1
 CC core protein, P24H, linked to a B-cell epitope, PRI which represents
 CC a consensus sequence derived from New York and Amsterdam HIV-1 isolates.
 CC Chimeric peptides such as this, may be used in the production of HIV-1
 CC vaccines. These peptide sequences may also be used in the production of
 CC multimeric peptides in which the peptides are C-terminally modified by
 CC the addition of a lys residue which is modified on its epsilon amino
 CC acid to carry an additional copy of the peptide molecule. The linear
 CC and multimeric peptides may be used for the treatment of AIDS by acting
 CC to displace the binding of HIV virus to human or animal cells or by
 CC disturbing the 3D organisation of the virus.

XX Sequence 33 AA;

Query Match 22.7%; Score 43.5; DB 16; Length 33;

Best Local Similarity 50.0%; Pred. No. 68;

Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 8 TLPGGPGPEFKTT-LRVONTL 26

DB 5 SIPIGPGRAFTTGPVONI 24

Search completed: July 16, 2003, 08:22:45
 Job time : 69 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:24:44 ; Search time 51 Seconds
(without alignments)
81.502 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192
Sequence: 1 YEHPHYGLPVGPGEFFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 113972

Minimum DB seq length: 0
Maximum DB seq length: 35

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pdb.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
- 5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pdb.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pdb.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pdb.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	21.9	29	9	US-10-106-698-7462
2	39	20.3	20	9	US-09-864-761-39770
3	39	20.3	21	10	US-09-864-761-46555
4	39	20.3	23	9	US-09-843-676-208
5	39	20.3	23	9	US-09-438-486-208
6	39	20.3	23	9	US-10-053-758-208
7	39	20.3	23	9	US-10-054-295-208
8	39	20.3	23	9	US-10-054-611-208
9	39	20.3	24	9	US-09-798-889-167
10	38	19.8	15	10	US-09-810-310-27
11	38	19.8	24	9	US-10-023-282-472
12	38	19.8	29	9	US-10-045-465-12
13	38	19.8	33	9	US-10-272-339A-18
14	38	19.8	33	9	US-10-272-339A-20
15	36.5	19.0	28	9	US-09-892-877-293
16	36.5	19.0	28	9	US-09-948-783-305
17	36	18.8	15	10	US-09-810-310-28
18	36	18.8	17	9	US-09-964-201A-3
19	36	18.8	22	9	US-09-932-613-167

20	36	18.8	25	9	US-10-050-882-98	Sequence 98, Appl
21	36	18.8	30	9	US-10-106-698-7473	Sequence 7473, Ap
22	36	18.8	33	9	US-10-106-698-7486	Sequence 7486, Ap
23	35.5	18.5	21	9	US-10-097-065-606	Sequence 606, App
24	35.5	18.5	34	9	US-10-106-698-7549	Sequence 7549, Ap
25	35	18.2	12	9	US-10-254-446A-193	Sequence 193, App
26	35	18.2	15	10	US-09-767-460-7	Sequence 7, Appl
27	35	18.2	15	10	US-09-810-310-30	Sequence 30, Appl
28	35	18.2	25	10	US-09-803-165-17	Sequence 17, Appl
29	35	18.2	28	10	US-09-864-761-4153	Sequence 4153, A
30	35	18.2	29	10	US-09-864-761-35300	Sequence 35300, A
31	35	18.2	31	10	US-09-864-761-43565	Sequence 43565, A
32	35	18.2	33	10	US-09-864-761-44246	Sequence 44246, A
33	34.5	18.0	21	9	US-10-106-698-8032	Sequence 8032, Ap
34	34.5	18.0	23	9	US-10-106-698-8221	Sequence 8221, Ap
35	34	17.7	10	9	US-09-087-513-18	Sequence 18, Appl
36	34	17.7	10	9	US-09-810-310-20	Sequence 20, Appl
37	34	17.7	17	9	US-10-059-271-18	Sequence 18, Appl
38	34	17.7	18	10	US-09-030-619-126	Sequence 126, App
39	34	17.7	19	9	US-10-059-271-33	Sequence 33, Appl
40	34	17.7	19	10	US-09-308-511-1	Sequence 1, Appl
41	34	17.7	20	9	US-10-062-831-203	Sequence 203, App
42	34	17.7	21	10	US-09-853-830-49	Sequence 49, Appl
43	34	17.7	28	9	US-09-462-713-7	Sequence 7, Appl
44	34	17.7	30	9	US-10-040-862-10444	Sequence 10444, A
45	34	17.7	30	9	US-10-023-282-1093	Sequence 1093, Ap

ALIGNMENTS

RESULT 1
US-10-106-698-7462
Sequence 7462, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:
APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
FILE REFERENCE: PA005051
CURRENT APPLICATION NUMBER: US/10/106,698
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: PCT/US00/26524
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/157,137
PRIOR FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: US 60/163,280
PRIOR FILING DATE: 1999-11-03
NUMBER OF SEQ ID NOS: 8564
SOFTWARE: PatentIn Ver. 3.0
SEQ ID NO 7462
LENGTH: 29
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (3)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids.
NAME/KEY: MISC FEATURE
LOCATION: (6)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (7)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: MISC FEATURE
LOCATION: (9)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-7462

Query Match
Best Local Similarity 56.2%; Pred. No. 45;
Matches 9; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 4 HPYGTLL--PVGPGEFF 17

Db 11 HYFGKLVPPAGTQPEF 26

RESULT 2

US-09-864-761-39770
Sequence 39770, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 39770
LENGTH: 20
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ACO05747.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.72
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.79
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EST_HUMAN HIT: H87835.1, EVALUATE 2.00e-06
US-09-864-761-39770

Query Match 20.3%; Score 39; DB 10; Length 20;

Best Local Similarity 58.3%; Pred. No. 77;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RESULT 3

US-09-864-761-46555
Sequence 46555, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 46555
LENGTH: 21
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ACO09973.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELLO, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
OTHER INFORMATION: EST_HUMAN HIT: AL121516.1, EVALUATE 8.00e-07

US-09-864-761-46555

Query Match	20.3%	Score 39;	DB 10;	Length 21;
Best Local Similarity	57.9%;	Pred. No. 81;		
Matches 11; Conservative	1;	Mismatches 3;	Indels 4;	Gaps 2

Db 6 DVH-RGTL--PGPNHKTT 20

RESULT 4
US-09-843-676-208

APPLICANT: Cech, Thomas R.

TITLE OF INVENTION: No. US20020164786A1el Telomerase
NUMBER OF SEQUENCES: 225

ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/843,676
FILING DATE: 26-Apr-2001
CLASSIFICATION: 536

APPLICATION NUMBER: US/08/854,050
FILING DATE: 09-MAY-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid

SEQUENCE	DESCRIPTION:	SEQ ID NO:	208:
US-09-843-676-208			

Query Match	20.3%	Score 39;	DB 9;	Length 23;
Best Local Similarity	58.3%;	Pred. No. 90;		
Matches 7;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0

Db 9 HPRENL PQDPGP 20

RESULT 5
US-09-438-486-208

APPLICANT: Linger, Joachim
 APPLICANT: Nakamura, Toru
 APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Harley, Calvin
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: No. US20030090919A1 Telomerase
 NUMBER OF SEQUENCES: 223

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

APPLICATION NUMBER: US/09/438,486
FILING DATE: 12-NOV-1999
CLASSIFICATION: 536

APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
CLASSIFICATION: 536

APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
CLASSIFICATION: 536

APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,663
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002931US
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 20

LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS:

US-09-438-486-208

Query Match	20.3%	Score 39;	DB 9;	Length 23;
Best Local Similarity	58.3%	Pred. No. 90;		
Matches	7; Conservative	0; Mismatches	5; Indels	0; Gaps

QY 4 HPYGTLPVGPGR 15

RESULT 6
US-10-053-758-208
; Sequence 208, Application US/10053758
; Publication No. US20030032075A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; ; Lingner, Joachim
; ; Nakamura, Toru
; ; Chapman, Karen B.
; ; Morin, Gregg B.
; ; Harley, Calvin
; ; Andrews, William H.
; TITLE OF INVENTION: No. US20030032075A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/053,758
; FILING DATE: 18-Jan-2002
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/854,050
; FILING DATE: 09-MAY-1997
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-10-053-758-208
Query Match 20.3%; Score 39; DB 9; Length 23;
Best Local Similarity 58.3%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 HPGTLPVGP 15
DB 9 HRENLPDPP 20
RESULT 7
US-10-054-295-208
; Sequence 208, Application US/10054295
; Publication No. US20030044953A1
; GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.
; ; Lingner, Joachim
; ; Nakamura, Toru
; ; Chapman, Karen B.
; ; Morin, Gregg B.
; ; Harley, Calvin
; ; Andrews, William H.
; TITLE OF INVENTION: No. US20030044953A1el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/054,295
; FILING DATE: 18-Jan-2002
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/854,050
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-002930US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-10-054-295-208
Query Match 20.3%; Score 39; DB 9; Length 23;
Best Local Similarity 58.3%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 HPGTLPVGP 15
DB 9 HRENLPDPP 20
RESULT 8
US-10-054-611-208
; Sequence 208, Application US/10054611
; Publication No. US20030059787A1
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; ; Lingner, Joachim
; ; Nakamura, Toru
; ; Chapman, Karen B.
; ; Morin, Gregg B.
; ; Harley, Calvin
; ; Andrews, William H.
; TITLE OF INVENTION: No. US20030059787A1el Telomerase

NUMBER OF SEQUENCES: 225
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/054,611
FILING DATE: 18-Jan-2002
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/854,050
FILING DATE: <Unknown>
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002930US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 208:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-10-054-611-208
Query Match 20.3%; Score 39; DB 9; Length 23;
Best Local Similarity 58.3%; Pred. No. 90;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 4 HRYGLPVGPGP 15
DB 9 HRENLPQDPGP 20
RESULT 9
US-09-798-889-167
Sequence 167, Application US/09798889
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 31 Human secreted proteins
FILE REFERENCE: P2026P1
CURRENT APPLICATION NUMBER: US/09/798,889
PRIOR FILING DATE: 2001-03-06
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022
PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,656
PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 185

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 167
LENGTH: 24
TYPE: PRT
ORGANISM: Homo sapiens
US-09-798-889-167
Query Match 20.3%; Score 39; DB 9; Length 24;
Best Local Similarity 54.5%; Pred. No. 94;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 YEHPYGLPV 11
DB 7 YTKHPFGHPV 17
RESULT 10
US-09-810-310-27
Sequence 27, Application US/09810310
Patent No. US20020044948A1
GENERAL INFORMATION:
APPLICANT: Khleif, Samir N.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR CO-STIMULATION OF
FILE REFERENCE: 15280-41510US
CURRENT APPLICATION NUMBER: US/09/810,310
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/189,396
PRIOR FILING DATE: 2000-03-15
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 27
LENGTH: 15
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: HIV PEPTIDE
US-09-810-310-27
Query Match 19.8%; Score 38; DB 10; Length 15;
Best Local Similarity 46.2%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 8 TLVPGPGPEFKT 20
DB 1 SIHIGPGPAFYAT 13
RESULT 11
US-10-023-282-472
Sequence 472, Application US/10023282
Publication No. US20030092893A1
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: P2007P1
CURRENT APPLICATION NUMBER: US/10/023,282
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 09/205,258
PRIOR FILING DATE: 1998-12-04
PRIOR APPLICATION NUMBER: PCT/US98/11422
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/048,885
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/049,375
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/048,881
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/048,880
PRIOR FILING DATE: 1997-06-06
PRIOR APPLICATION NUMBER: 60/048,896

```

1 EARLIER FILING DATE: 1997-06-06
2 EARLIER APPLICATION NUMBER: 60/049, 020
3 EARLIER FILING DATE: 1997-06-06
4 EARLIER APPLICATION NUMBER: 60/048, 876
5 EARLIER FILING DATE: 1997-06-06
6 EARLIER APPLICATION NUMBER: 60/048, 8935
7 EARLIER FILING DATE: 1997-06-06
8 EARLIER APPLICATION NUMBER: 60/048, 8933
9 EARLIER FILING DATE: 1997-06-06
10 EARLIER APPLICATION NUMBER: 60/048, 884
11 EARLIER APPLICATION NUMBER: 60/048, 894
12 EARLIER FILING DATE: 1997-06-06
13 EARLIER APPLICATION NUMBER: 60/048, 9711
14 EARLIER FILING DATE: 1997-06-06
15 EARLIER APPLICATION NUMBER: 60/048, 882
16 EARLIER FILING DATE: 1997-06-06
17 EARLIER APPLICATION NUMBER: 60/048, 899
18 EARLIER FILING DATE: 1997-06-06
19 EARLIER APPLICATION NUMBER: 60/048, 893
20 EARLIER FILING DATE: 1997-06-06
21 EARLIER APPLICATION NUMBER: 60/048, 900
22 EARLIER FILING DATE: 1997-06-06
23 EARLIER APPLICATION NUMBER: 60/048, 901
24 EARLIER FILING DATE: 1997-06-06
25 EARLIER APPLICATION NUMBER: 60/048, 892
26 EARLIER FILING DATE: 1997-06-06
27 EARLIER APPLICATION NUMBER: 60/048, 915
28 EARLIER FILING DATE: 1997-06-06
29 EARLIER APPLICATION NUMBER: 60/049, 019
30 EARLIER FILING DATE: 1997-06-06
31 EARLIER APPLICATION NUMBER: 60/048, 970
32 EARLIER FILING DATE: 1997-06-06
33 EARLIER APPLICATION NUMBER: 60/048, 9722
34 EARLIER FILING DATE: 1997-06-06
35 EARLIER APPLICATION NUMBER: 60/048, 916
36 EARLIER FILING DATE: 1997-06-06
37 EARLIER APPLICATION NUMBER: 60/049, 373
38 EARLIER FILING DATE: 1997-06-06
39 EARLIER APPLICATION NUMBER: 60/048, 875
40 EARLIER FILING DATE: 1997-06-06
41 EARLIER APPLICATION NUMBER: 60/049, 374
42 EARLIER FILING DATE: 1997-06-06
43 EARLIER APPLICATION NUMBER: 60/048, 917
44 EARLIER FILING DATE: 1997-06-06
45 EARLIER APPLICATION NUMBER: 60/048, 949
46 EARLIER FILING DATE: 1997-06-06
47 EARLIER APPLICATION NUMBER: 60/048, 974
48 EARLIER FILING DATE: 1997-06-06
49 EARLIER APPLICATION NUMBER: 60/048, 883
50 EARLIER FILING DATE: 1997-06-06
51 EARLIER APPLICATION NUMBER: 60/048, 897
52 EARLIER FILING DATE: 1997-06-06
53 EARLIER APPLICATION NUMBER: 60/048, 898
54 EARLIER FILING DATE: 1997-06-06
55 EARLIER APPLICATION NUMBER: 60/048, 962
56 EARLIER FILING DATE: 1997-06-06
57 EARLIER APPLICATION NUMBER: 60/048, 963
58 EARLIER FILING DATE: 1997-06-06
59 EARLIER APPLICATION NUMBER: 60/048, 877
60 EARLIER FILING DATE: 1997-06-06
61 EARLIER APPLICATION NUMBER: 60/048, 878
62 EARLIER FILING DATE: 1997-12-18
63 EARLIER APPLICATION NUMBER: 60/092, 921
64 EARLIER FILING DATE: 1998-07-15
65 EARLIER APPLICATION NUMBER: 60/094, 657
66 EARLIER FILING DATE: 1998-07-30
67 NUMBER OF SEQ ID NOS: 1227
68 SOFTWARE: PatentIn Ver. 2.0
69 SEQ ID NO 472
70 LENGTH: 24

```

```

; TYPE: ERT
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: SITE
; LOCATION: (24)
; OTHER INFORMATION: Xaa equals stop translation
;S-10-023-262-472
```

Query Match	19.8%	Score 38;	DB 9;	Length 24;
Best Local Similarity	66.7%	Pred. No. 1.3e+02;		
Matches	8;	Conservative	0;	Mismatches 2;
			Indels	2;
			Gaps	1;

QY 4 HPYGTLPVGPGR 15
|||
Db 13 HPPG--PQGPGR 22

RESULT 12
US-10-045-465-12
; Sequence 12, Application US/10045465
; Publication No. US20030049605A1
GENERAL INFORMATION

```

1 GENERAL INFORMATION:
2 APPLICANT: Van Es, Helmut
3 TITLE OF INVENTION: Display of viral proteins
4 FILE REFERENCE: 2183-4079US
5 CURRENT APPLICATION NUMBER: US/10/045,465
6 PRIOR FILING DATE: 2001-11-07
7 PRIOR APPLICATION NUMBER: 09/315,244
8 PRIOR FILING DATE: 1999-05-20
9 PRIOR APPLICATION NUMBER: EP 98201678.4
10 PRIOR FILING DATE: 1998-05-20
11 NUMBER OF SEQ ID NOS: 15
12 SOFTWARE: Patentin version 3.0
13 SEQ ID NO 12

```

```

? LENGTH: 29
? TYPE: PR1
? ORGANISM: Moloney murine leukemia virus
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (1)..(29)
? OTHER INFORMATION: /No. US20030049605A1e="Amino acid sequence of ecotropic env varia
? OTHER INFORMATION: and/or parts:
? OTHER INFORMATION: COOH-terminus of construct 3 at location 251-266
? US-10-045-465-12

```

Query Match	19.8%	Score 38;	DB 9;	Length 29;
Best Local Similarity	69.2%	Pred. No. 1.6e+02;		
Matches	9;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0;

QY	21	LRVQNLGCVVSG	33
Db	4	LRVQNLGPRVTS	16

RESULT 13
US-10-272-339A-18

TITLE OF INVENTION: T Cell Receptor VB-DB-JB Sequence and Methods For Its
 TITLE OF INVENTION: Detection
 FILE REFERENCE: 213838-00029
 CURRENT APPLICATION NUMBER: US/10/272,339A
 CURRENT FILING DATE: 2002-10-16
 PRIOR APPLICATION NUMBER: US 09/641,576
 PRIOR FILING DATE: 2000-08-18
 PRIOR APPLICATION NUMBER: US 09/507,819
 PRIOR FILING DATE: 2000-02-22
 PRIOR APPLICATION NUMBER: US 60/121,311
 PRIOR FILING DATE: 1999-02-23
 NUMBER OF SEQ ID NOS: 77
 SOFTWARE: Patentlin Ver. 2.1

```

; SEQ ID NO 18
; LENGTH: 33
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-272-339A-18

```

Query Match	19.8%	Score 38	DB 9	Length 33
Best Local Similarity	40.0%	Pred. No. 1.9e+02		
Matches 10; Conservative	2	Mismatches 11	Indels 2	Gaps 1

QY 3 VHPYG--TLPPVGPGPEFKITLRVQ 25
| | | | : | : ||
Db 9 VRPGGNKLTFTGTQLKVELINIQ 33

```

RESULT 14
US-10-272-339A-20
: Sequence 20, Application US/10272339A
: Publication No. US20030120061A1
: GENERAL INFORMATION:
: APPLICANT: Jingwu, Zhang Z.
: TITLE OF INVENTION: T Cell Receptor VB-DB-JB Sequence and Methods For Its
: TITLE OF INVENTION: Detection
: FILE REFERENCE: 213838-00029
: CURRENT APPLICATION NUMBER: US/10/272,339A
: CURRENT FILING DATE: 2002-10-16
: PRIOR APPLICATION NUMBER: US 09/641,576
: PRIOR FILING DATE: 2000-08-18
: PRIOR APPLICATION NUMBER: US 09/507,819
: PRIOR FILING DATE: 2000-02-22
: PRIOR APPLICATION NUMBER: US 60/121,311
: PRIOR FILING DATE: 1999-02-23
: NUMBER OF SEQ ID NOS: 77
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 20
: LENGTH: 33
: TYPE: prt
: ORGANISM: Homo sapiens
: US-10-272-339A-20

```

Query Match	19.8%	Score 38;	DB 9;	Length 33;
Best Local Similarity	40.0%;	Pred. No. 1.9e+02;		
Matches 10; Conservative	2;	Mismatches 11;	Indels 2;	Gaps 1;

QY 3 VHPYG--TLPVGPGPEFKITLRVQN 25
| | | | : | | : ||
DB 9 VRPGGKLTFTGTQTLKVELNIQN 33

```

RESULT 15
US-09-892-877-293
; Sequence 293, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 293
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-892-877-293

```

Query Match	19.0%	Score 36.5	DB 9	length 28
Best Local Similarity	72.7%	Pred. NO. 2.5e+02		
Matches	8	Conservative	0	Mismatches 2; Indels 1; Gaps 1;

Qy	7	GTLPVGPGEF	17
Db	1	GT-PAGTGPEF	10

Search completed: July 16, 2003, 08:31:21
Job time : 53 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:21:34 ; Search time 26 Seconds
(without alignments)
39,608 Million cell updates/sec

Title: US-09-647-544-2_COPY_952_986

Perfect score: 192

Sequence: 1 YEHPYGTLPVGPPEFKTLRVQNLGCVVSGLI 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 167989

Minimum DB seq length: 0
Maximum DB seq length: 35

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA: +
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	24.0	33	1	US-08-257-528B-84
2	46	24.0	33	1	US-08-460-602A-84
3	46	24.0	33	1	US-08-463-966A-84
4	46	24.0	33	1	US-08-465-217A-84
5	46	24.0	33	2	US-08-464-329A-84
6	46	24.0	33	2	US-08-462-507A-84
7	46	24.0	33	2	US-08-467-881A-84
8	46	24.0	34	4	US-09-141-833-11
9	45.5	23.7	33	1	US-08-257-528B-83
10	45.5	23.7	33	1	US-08-460-602A-83
11	45.5	23.7	33	1	US-08-463-966A-83
12	45.5	23.7	33	1	US-08-465-217A-83
13	45.5	23.7	33	2	US-08-464-329A-83
14	45.5	23.7	33	2	US-08-462-507A-83
15	45.5	23.7	33	2	US-08-467-881A-83
16	45	23.4	35	3	US-08-513-968-13
17	43.5	22.7	33	1	US-08-257-528B-81
18	43.5	22.7	33	1	US-08-460-602A-81
19	43.5	22.7	33	1	US-08-463-966A-81
20	43.5	22.7	33	1	US-08-465-217A-81
21	43.5	22.7	33	2	US-08-464-329A-81
22	43.5	22.7	33	2	US-08-462-507A-81
23	43.5	22.7	33	2	US-08-467-881A-81
24	43	22.4	18	1	US-08-257-528B-50
25	43	22.4	18	1	US-08-460-602A-50
26	43	22.4	18	1	US-08-463-966A-50
27	43	22.4	18	1	US-08-465-217A-50

28	43	22.4	18	2	US-08-464-329A-50	Sequence 50, Appl
29	43	22.4	18	2	US-08-462-507A-50	Sequence 50, Appl
30	43	22.4	18	2	US-08-467-881A-50	Sequence 50, Appl
31	43	22.4	23	3	US-08-513-968-71	Sequence 71, Appl
32	43	22.4	23	3	US-08-513-968-72	Sequence 72, Appl
33	43	22.4	23	1	US-08-257-528B-41	Sequence 41, Appl
34	43	22.4	33	1	US-08-257-528B-80	Sequence 41, Appl
35	43	22.4	33	1	US-08-460-602A-80	Sequence 41, Appl
36	43	22.4	33	1	US-08-463-966A-80	Sequence 41, Appl
37	43	22.4	33	1	US-08-465-217A-80	Sequence 41, Appl
38	43	22.4	33	1	US-08-467-881A-80	Sequence 41, Appl
39	43	22.4	33	1	US-08-465-217A-41	Sequence 41, Appl
40	43	22.4	33	1	US-08-465-217A-80	Sequence 41, Appl
41	43	22.4	33	2	US-08-464-329A-41	Sequence 41, Appl
42	43	22.4	33	2	US-08-464-329A-80	Sequence 41, Appl
43	43	22.4	33	2	US-08-462-507A-41	Sequence 41, Appl
44	43	22.4	33	2	US-08-462-507A-80	Sequence 41, Appl
45	43	22.4	33	2	US-08-467-881A-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-08-257-528B-84
Sequence 84, Application US/08257528B
Patent No. 5639854
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,528B
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-336 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-528B-84

Query Match 24.0%; Score 46; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

Cy 8 TLVPGGPPEFKT-----LRVQNLG 27
Db 5 SIPIGGRAPFTTGYKRVVKIPIPG 30

```
RESULT 2
US-08-460-602A-84
; Sequence 84, Application US/08460602A
; Patent No. 5759769
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,602A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-450 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-460-602A-84
;
Query Match      24.0%; Score 46; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLPGVGPGEFKT-----LRVQNLG 27
Db      5 SIPIGGRGYTTGKYKVVKIEPLG 30

RESULT 3
US-08-463-966A-84
; Sequence 84, Application US/08463966A
; Patent No. 5795955
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
```

```

; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,966A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-487 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-463-966A-84
;
Query Match      24.0%; Score 46; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLPGVGPGEFKT-----LRVQNLG 27
Db      5 SIPIGGRGYTTGKYKVVKIEPLG 30

RESULT 4
US-08-465-217A-84
; Sequence 84, Application US/08465217A
; Patent No. 5800822
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,217A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
```



```
; Sequence 84, Application US/08467881A
; Patent No. 5951986
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,881A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,528
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/073,378
; FILING DATE: 09-JUN-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-488 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-467-881A-84

Query Match      24.0%; Score 46; DB 2; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.1;
Matches 9; Conservative 7; Mismatches 4; Indels 6; Gaps 1;

QY      8 TLVPGGPEFKT-----LRVQNLG 27
DB      5 SLIPGGRAPFTTGYKVKVKKLEPLG 30

RESULT 8
US-09-141-833-11
; Sequence 11, Application US/09141833
; Patent No. 6168784
; GENERAL INFORMATION:
; APPLICANT: OFFORD, ROBIN E
; APPLICANT: THOMPSON, DAREN
; APPLICANT: WILKEN, JILL
; TITLE OF INVENTION: N-TERMINAL MODIFICATIONS OF RANTES AND METHODS OF USE
; FILE REFERENCE: GREN-026/03US
; CURRENT APPLICATION NUMBER: US/09/141,833
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: 60/056,292
; EARLIER FILING DATE: 1997-09-03
; EARLIER APPLICATION NUMBER: 60/077,874
; EARLIER FILING DATE: 1998-03-13
; EARLIER APPLICATION NUMBER: 60/090,834
```

```
; EARLIER FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO. 11
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-141-833-11

Query Match      24.0%; Score 46; DB 4; Length 34;
Best Local Similarity 47.4%; Pred. No. 4.3;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      8 TLVPGGPEFKTTLRVQNL 26
DB      11 SLIPGGRAPFTTGYKVKVKKLEPLG 29

RESULT 9
US-08-257-528B-83
; Sequence 83, Application US/08257528B
; Patent No. 5639854
; GENERAL INFORMATION:
; APPLICANT: SIA, Charles D.Y.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,528B
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, MICHAEL I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-336 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-528B-83

Query Match      23.7%; Score 45.5; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY      2 EVHPYG-----TLVPGGPEFKT 20
DB      7 KIEPLGVAENPKRSIPGGRAPFTT 32

RESULT 10
US-08-460-602A-83
; Sequence 83, Application US/08460602A
; Patent No. 5759769
; GENERAL INFORMATION:
```


APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,602A
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-450 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-460-602A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

Qy 2 EVHPYG-----TLPVGSGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 11
US-08-463-966A-83
Sequence 83, Application US/08463966A
Patent No. 5795955
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,966A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-487 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-463-966A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

Qy 2 EVHPYG-----TLPVGSGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

RESULT 12
US-08-465-217A-83
Sequence 83, Application US/08465217A
Patent No. 5800822
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,217A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-486 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-465-217A-83

Query Match 23.7%; Score 45.5; DB 1; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPVGPGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGPGRAFYTT 32

RESULT 13
US-08-464-329A-83
Sequence 83, Application US/08464329A
Patent No. 5817754
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,329A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-449 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-464-329A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;

Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPVGPGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGPGRAFYTT 32

RESULT 14
US-08-462-507A-83
Sequence 83, Application US/08462507A
Patent No. 5876731
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,507A
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-451 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-462-507A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 EVHPYG-----TLVPVGPGPEPKTT 20
Db 7 KIEPLGVAPNTRKSIPIGPGRAFYTT 32

RESULT 15
US-08-467-881A-83
Sequence 83, Application US/08467881A
Patent No. 5851986
GENERAL INFORMATION:
APPLICANT: SIA, Charles D.Y.
APPLICANT: CHONG, Pele

APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Tandem Synthetic HIV-1 Peptides
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,881A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,528
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/073,378
FILING DATE: 09-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, MICHAEL I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-488 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-467-881A-83

Query Match 23.7%; Score 45.5; DB 2; Length 33;
Best Local Similarity 34.6%; Pred. No. 4.9;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;
QY 2 EVHPYG-----TLPVGGPGEFFKT 20
Db 7 KIEPLGVAPNTRKSIPIGGRAFYTT 32

Search completed: July 16, 2003, 08:25:59
Job time : 27 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:52 ; Search time 1.1294 Seconds

(without alignments)
1872.646 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114
Sequence: 1 KLGFPAHKKIPEEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seque, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	46.5	214	2 A11073	probable membrane
2	51	44.7	672	2 G69503	signal-transducing
3	51	44.7	856	2 H64552	endopeptidase Clp
4	51	44.7	1045	2 S60571	integrin alpha v c
5	50	43.9	357	2 AC1104	B. subtilis YacI p
6	50	43.9	505	1 S77034	protein kinase pkn
7	50	43.9	1034	2 A36108	integrin alpha-v c
8	50	43.9	1044	2 T10050	integrin alpha-v c
9	50	43.9	1048	2 A27421	integrin alpha-5 c
10	49.5	43.4	236	2 T12766	probable lipoprote
11	49.5	43.4	959	1 B60017	outer capsid prote
12	49	43.0	357	2 AC1466	B. subtilis YacI p
13	49	43.0	655	2 T06108	hypothetical prote
14	49	43.0	853	2 S74279	hypothetical prote
15	49	43.0	856	2 C71856	probable endopepti
16	48.5	42.5	407	2 S23325	M2 protein precurs
17	48	42.1	421	2 H90433	hypothetical prote
18	48	42.1	573	2 A55767	squalene monooxyge
19	48	42.1	962	2 JCS5808	G protein-coupled
20	48	42.1	1464	2 T13716	bazooka gene prote
21	48	42.1	1526	2 JN0598	DNA topoisomerase
22	47	41.2	102	2 F75077	hypothetical prote
23	47	41.2	112	2 E70433	flagellar switch p
24	47	41.2	206	2 G97612	hypothetical prote
25	47	41.2	512	2 G98236	probable transposa
26	47	41.2	512	2 AC1303	IS3 family transpo
27	47	41.2	512	2 AD2835	IS3 family transpo
28	47	41.2	512	2 AD3049	IS3 family transpo
29	47	41.2	512	2 AD3049	IS3 family transpo

30	47	41.2	692	2 S37976	hypothetical prote
31	47	41.2	861	2 T02267	trehalose-6-phosph
32	46	40.4	278	2 D84492	hypothetical prote
33	46	40.4	308	2 S11553	oligopeptide trans
34	46	40.4	308	2 E95220	hypothetical prote
35	46	40.4	323	2 E98084	hypothetical prote
36	46	40.4	450	2 S37900	hypothetical prote
37	46	40.4	544	2 AH2971	hypothetical prote
38	46	40.4	544	2 B96311	probable ATP-bind
39	46	40.4	847	2 C96703	hypothetical prote
40	46	40.4	1037	2 A60163	glycoprotein fib
41	46	40.4	1407	2 S28589	trichonhyalin - tab
42	45.5	39.9	570	2 F70332	proline-trna synth
43	45	39.5	273	2 T47612	hypothetical prote
44	45	39.5	325	2 T18283	hypothetical prote
45	45	39.5	333	2 T08850	alternative respir

ALIGNMENTS

RESULT 1

A11073

probable membrane protein smp [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhimurium

A:Note: this species has also been called Salmonella typhimurium

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #ext_change 09-Nov-2001

C/Accession: A11073

R:Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., et al., 2001

Nature 413, 848-852, 2001

A:Authors: Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., et al., 2001

A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A/Reference number: AB0502; PMID:11677608

A/Accession: A11073

A>Status: preliminary

A/Molecule type: DNA

A/Residues: 1-214 <PAR>

A/Cross-references: GB:AL513382; PIDN:CMD03408.1; PID:g16505677; GSPDB:GN00176

A:Gene: smp

Query Match 46.5%; Score 53; DB 2; Length 214;

Best local similarity 71.4%; Pred. No. 5;

Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREK 19

DB 200 ASKPVEPEERREK 213

RESULT 2

G69503

signal-transducing histidine kinase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #ext_change 21-Jul-2000

C/Accession: G69503

R:Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, J., et al., 1997

Nature 390, 364-370, 1997

A:Authors: Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, J., et al., 1997

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus

A/Reference number: A69250; PMID:98049343; PMID:9389475

A/Accession: G69503

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-672 <KLE>

A/Cross-references: GB:AE000963; GB:AE000782; NID:g2689286; PIDN:AA89224.1; PID:g26485C

Query Match 44.7%; Score 51; DB 2; Length 672;

Best Local Similarity 55.6%; Pred. No. 30;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 5 FAHKKIPBEKREKLEO 22
DB 572 FAFKSMEDEERREELKQ 589

RESULT 3

H64552

endopeptidase Clp ATP-binding chain B - Helicobacter pylori (strain 26695)
N/Alternate names: ATP-dependent Clp proteinase regulatory chain
N/Contains: adenosinetriphosphatase (EC 3.6.1.3)
C/Species: Helicobacter pylori

C/Date: 09-Aug-1997 #sequence_revision 15-Aug-1997 #text_change 19-Jan-2001

C/Accession: H64552

R/Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Kalak, H.G.; Glodek, A.; McKen-

son, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman, J.M.; Fujii, C.; Bowman, C.; Wathey, L.

Nature 388, 539-547, 1997

A/Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A/Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A/Reference number: A64520; PMID:97394467; PMID:9252185

A/Accession: H64552

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-856 <TOM>

A/Cross-references: GB:AE000545; GB:AE000511; NID:92313349; PIDN:MAD07330.1; PID:g21335

C/Function: allows clp to hydrolyze polypeptides and proteins, probably by a chaperon

e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller

C/Superfamily: endopeptidase Clp ATP-binding chain

C/Keywords: ATP; duplication; hydrolase; molecular chaperone; nucleotide binding; P-loop

P:204-211/Region: nucleotide-binding motif A (P-loop)

F:211-276/Region: nucleotide-binding motif B

F:605-612/Region: nucleotide-binding motif A (P-loop)

F:673-678/Region: nucleotide-binding motif B

F:210/Binding site: ATP (Lys) #status predicted

F:611/Binding site: ATP (Lys) #status predicted

Query Match 44.7%; Score 51; DB 2; Length 856;

Best Local Similarity 45.0%; Pred. No. 38;

Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 3 GFPAKKIPBEKREKLEO 22

DB 496 GEIRYSKIPENKKEELOR 515

RESULT 4

S60571

integrin alpha v chain precursor - Iberian ribbed newt

C/Species: Pleurodeles waltl (Iberian ribbed newt)

C/Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 29-Sep-1999

C/Accession: S60571

R/Alfandari, D.; Whitaker, C.A.; Desimone, D.W.; Daribere, T.

Dev. Biol. 170, 249-261, 1995

A/Title: Integrin alpha-(v) subunit is expressed on mesodermal cell surfaces during amphi-

A/Reference number: S60571; PMID:95377519; PMID:7649360

A/Accession: S60571

A/Status: preliminary; nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-1045 <RLF>

A/Cross-references: EMBL:X81108; NID:g1008137; PIDN:CA57014.1; PID:g1008138

C/Superfamily: integrin alpha-2b chain

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-1045/Product: integrin alpha v chain #status predicted <MAT>

Query Match 44.7%; Score 51; DB 2; Length 1045;

Best Local Similarity 42.9%; Pred. No. 47;

Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFAKKIPBEKREKLE 21

DB 1014 KFGFKVRPQGEGEREQLQ 1034

RESULT 5

AC1104

B. subtilis yacI protein homolog lmo0234 [imported] - Listeria monocytogenes (strain EGI

C/Species: Listeria monocytogenes

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C/Accession: AC1104

R/Glaeser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baguer, F.; Berche, P.; Bloeker,

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussange, O.; Entian, K.D.; Fsihl, H

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madueno, E.; Maitournam, A.; M-

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland,

A/Title: Comparative genomics of Listeria species.

A/Reference number: AB1077; PMID:21537279; PMID:11679669

A/Accession: AC1104

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1357 <GLA>

A/Cross-references: GB:NC_003210; PIDN:CAD00761.1; PID:g16409599; GSPDB:GN00177

A/Experimental source: strain EGD-e

C/Genetics:

A/Gene: lmo0234

C/Superfamily: conserved hypothetical protein yacI

Query Match 43.9%; Score 50; DB 2; Length 357;

Best Local Similarity 71.4%; Pred. No. 22;

Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 AHKKIPBEKREK 19

DB 146 AKKRTPEKREK 159

RESULT 6

S77034

protein kinase pknA (EC 2.7.1.-), 55K - Synechocystis sp. (strain PCC 6803)

N/Alternate names: protein sl10776

C/Species: Synechocystis sp.

A/Variety: PCC 6803

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C/Accession: S77034

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asami, E.; Nakamura, Y.; Miyajima, N.

o, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud.

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti-

s.

A/Reference number: S74322; PMID:97061201; PMID:8905231

A/Accession: S77034

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-505 <KAN>

A/Cross-references: EMBL:D64005; GB:AB001339; NID:g1001779; PIDN:BA10726.1; PID:g10065

C/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Genetics:

A/Gene: pknA

C/Superfamily: Synechocystis sp. protein kinase pknA, 55K; protein kinase homology

C/Keywords: phosphotransferase; protein kinase

F:7-268/Domain: protein kinase homology <KIN>

Query Match 43.9%; Score 50; DB 1; Length 505;

Best Local Similarity 58.8%; Pred. No. 31;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 AHKKIPBEKREKLEO 22

DB 376 AEKLIANEKRORELEQ 392

RESULT 7

A36108
 Integrin alpha-V chain precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 29-Sep-1999
 C:Accession: A36108
 R:Bossy, B.; Reichardt, L.F.
 Biochemistry 29, 10191-10198, 1990
 A:Title: Chick integrin alpha-v subunit molecular analysis reveals high conservation of
 A:Reference number: A36108; MUID:91104936; PMID:1703004
 A:Accession: A36108
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1034 <BOS>
 A:Cross-references: GB:M60517; NID:G212876; PIDN:AAA49138.1; PID:G212877
 C:Superfamily: Integrin alpha-2b chain
 C:Keywords: cell adhesion; cytoskeleton; transmembrane protein

Query Match 43.9%; Score 50; DB 2; Length 1034;
 Best Local Similarity 38.1%; Pred. No. 64;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21
 1003 RMGFFKVRPPQEEQEREQLQ 1023

RESULT 8
 T10050
 Integrin alpha-v chain precursor - mouse
 N:Alternate names: vitronectin receptor alpha chain
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T10050
 R:Wada, J.; Kumar, A.; Liu, Z.; Ruoslahti, E.; Reichardt, L.; Marvaldi, J.; Kanwar, Y.S.
 J. Cell Biol. 133, 1161-1176, 1996
 A:Title: Cloning of mouse integrin alpha v cDNA and role of the alpha v-related matrix
 A:Reference number: Z16920; MUID:96176309; PMID:8601592
 A:Accession: T10050
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1044 <MAD>
 A:Cross-references: EMBL:U04135; NID:G537490; PIDN:MAC52497.1; PID:G537491
 A:Experimental source: strain CD-1, kidney
 C:Superfamily: Integrin alpha-2b chain
 C:Keywords: cell adhesion; receptor

Query Match 43.9%; Score 50; DB 2; Length 1044;
 Best Local Similarity 38.1%; Pred. No. 64;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21
 1013 RMGFFKVRPPQEEQEREQLQ 1033

RESULT 9
 A27421
 Integrin alpha-5 chain precursor - human
 N:Alternate names: CD51 antigen; serum spreading factor; vitronectin receptor alpha chain
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 31-Dec-2000
 C:Accession: A27421; S47541; B39418; A32287; A30298; A35035
 R:Suzuki, S.; Argaves, W.S.; Arai, H.; Languino, L.R.; Pierschbacher, M.D.; Ruoslahti, J.
 J. Biol. Chem. 262, 14080-14085, 1987
 A:Title: Amino acid sequence of the vitronectin receptor alpha subunit and comparative
 A:Reference number: A27421; MUID:98007656; PMID:2443500
 A:Accession: A27421
 A:Molecule type: mRNA
 A:Residues: 1-1048 <SUZ>
 A:Cross-references: GB:M14648; GB:J02826; GB:M18365; NID:G340306; PIDN:AAA36808.1; PID:G
 R:Donahue, J.P.; Sugg, N.; Hawiger, J.
 Biochim. Biophys. Acta 1219, 228-232, 1994
 A:Title: The integrin alpha(v) gene: identification and characterization of the promoted

A:Reference number: S47541; MUID:94368864; PMID:7522056
 A:Accession: S47541
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-61 <DON>
 A:Cross-references: EMBL:U07375
 R:Fitzgerald, L.A.; Poncz, M.; Steiner, B.; Rall Jr., S.C.; Bennett, J.S.; Phillips, D.
 Biochemistry 26, 8158-8165, 1987
 A:Title: Comparison of CDNA-derived protein sequences of the human fibronectin and vitr
 A:Reference number: A90526; MUID:88163472; PMID:2450560
 A:Accession: B29418
 A:Molecule type: mRNA
 A:Residues: 1-433 <FIT>
 R:Suzuki, S.; Argaves, W.S.; Pytela, R.; Arai, H.; Krusius, T.; Pierschbacher, M.D.; Ru
 Proc. Natl. Acad. Sci. U.S.A. 83, 8614-8618, 1986
 A:Title: cDNA and amino acid sequences of the cell adhesion protein receptor recognizing
 A:Reference number: A26482; MUID:87041504; PMID:2430295
 A:Accession: A26482
 A:Molecule type: mRNA
 A:Residues: 413-1048 <SUZ>
 R:Cheresh, D.A.; Smith, J.W.; Cooper, H.M.; Quaranta, V.
 Cell 57, 59-69, 1989
 A:Title: A novel vitronectin receptor integrin (alpha-v beta-x) is responsible for dist
 A:Reference number: A32287; MUID:89195223; PMID:2467745
 A:Accession: A32287
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 31-35, 'X', '37-41 <CHE>
 A:Experimental source: melanoma cell M21
 A:Accession: B32287
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 31-33, 'X', '35, 'X', '37-41 <CH2>
 A:Experimental source: lung carcinoma cell UCLA-P3
 R:Iam, S.C.T.; Plow, E.F.; D'Souza, S.E.; Cheresh, D.A.; Frelinger III, A.L.; Ginsberg,
 U. Biol. Chem. 264, 3742-3749, 1989
 A:Title: Isolation and characterization of a platelet membrane protein related to the v
 A:Reference number: A30298; MUID:89139425; PMID:2465293
 A:Accession: A30298
 A:Molecule type: protein
 A:Residues: 31-35, 'X', '37-40 <LAM>
 R:Smith, J.W.; Cheresh, D.A.
 J. Biol. Chem. 265, 2168-2172, 1990
 A:Title: Integrin (alpha-v beta-3)-ligand interaction. Identification of a heterodimeric
 A:Reference number: A35035; MUID:90130470; PMID:1688848
 A:Accession: A35035
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 66-72; 169-171, 'X', 173-176; 221-230; 255-258, 'X', 260; 325-328; 342-351; 466-473 <
 C:Genetics:
 A:Gene: GDB:ITGAV; VNRA
 A:Cross-references: GDB:120491; OMIM:193210
 A:Map position: 2q31-2q32
 C:Superfamily: Integrin alpha-2b chain
 C:Keywords: cell adhesion; duplication; glycoprotein; heterodimer; receptor; transmembr
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1048/Product: integrin alpha-v chain #status experimental <MAM>
 F:31-992/Domain: extracellular #status predicted <EXT>
 F:993-1016/Domain: transmembrane #status predicted <TMN>
 F:1017-1048/Domain: intracellular #status predicted <INT>

Query Match 43.9%; Score 50; DB 2; Length 1048;
 Best Local Similarity 38.1%; Pred. No. 65;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Db 1 KLGFFAKKIPPEEKREKLE 21
 1017 RMGFFKVRPPQEEQEREQLQ 1037

RESULT 10
 T12766
 Probable lipoprotein yolk - Bacillus subtilis phage SPBC2

C:Species: *Bacillus subtilis* phage SPBc2
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
 C:Accession: T12766; F63907
 R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Manuel, C.; Karamata, D.
 Submitted to the EMBL Data Library, August 1997
 A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBact2 prophage
 A:Reference number: 217583
 A:Accession: T12766
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-236 <L26>
 A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025480; PIDN:AC12975.1
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Aevedo, V.; Berth
 C.: Bron, S.; Brouillet, S.; Brunsch, C.V.; Caldwell, B.; Capuano, F.; Carter, N.M.; Ch
 A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, U.; Fabel, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall
 lech, J.; Harwood, C.R.; Hentzel, A.; Hilbert, H.; Holappell, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Manes
 Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Saadate, Y.; Sato, T.; Scanlon,
 A:Authors: Schleif, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Serd
 Akench, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033; PMID:9354377
 A:Accession: F63907
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-236 <KUN>
 A:Cross-references: GB:299115; GB:AL009126; NID:g2634478; PIDN:CAM14083.1; PID:el18312;
 A:Experimental source: strain 168
 A:Genetics:
 A:Gene: yokB

Query Match 43.4%; Score 49.5; DB 2; Length 236;
 Best Local Similarity 84.6%; Pred. No. 17;
 Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 8 KKI-PEEEKKEEK 19
 |||:|||||
 Db 205 KKVSPPEEKKEEK 217

RESULT 11
 B60017
 Outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)
 C:Species: bluetongue virus
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
 C:Accession: B60017
 R:Gould, A.R.; Pritchard, L.I.
 Virus Res. 17, 31-52, 1990
 A:Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and
 A:Reference number: A60017; MUID:91021485; PMID:2171239
 A:Accession: B60017
 A:Molecule type: genomic RNA
 A:Residues: 1-959 <GOU>
 A:Cross-references: GB:X55801; NID:g297130; PIDN:CAA39323.1; PID:g297131
 C:Genetics:
 A:Map position: segment 2
 C:Superfamily: bluetongue virus VP2 protein
 C:Keywords: capsid protein; glycoprotein
 F:749/910/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.4%; Score 49.5; DB 1; Length 959;
 Best Local Similarity 57.1%; Pred. No. 69;
 Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 2 LGFFAHHKKIPEEEKKEEK 22
 |||:|||||
 Db 791 LNFF-----PSYKREKKEE 806

RESULT 12
 AC1466
 B. subtilis YacI protein homolog lin0266 [imported] - *Listeria innocua* (strain Clp11262
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C:Accession: AC1466
 R:Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Faini, H
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefit, U.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Maguano, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of *Listeria* species
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1466
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-357 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CMC95499.1; PID:g16412695; GSPDB:GN00178
 A:Experimental source: strain Clp11262
 C:Genetics:
 C:Superfamily: conserved hypothetical protein yacI

Query Match 43.0%; Score 49; DB 2; Length 357;
 Best Local Similarity 71.4%; Pred. No. 30;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 AHKKIPEEEKKEEK 19
 |||:|||||
 Db 146 AKKKTPEEEKKEEK 159

RESULT 13
 T06108
 Hypothetical protein TSJ17.190 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999
 C:Accession: T06108
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.;
 Submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z15184
 A:Accession: T06108
 A:Molecule type: DNA
 A:Residues: 1-615 <BEV>
 A:Cross-references: EMBL:AL035708; GSPDB:GN00062; ATSP:TSJ17.190
 A:Experimental source: cultivar Columbia; BAC clone TSJ17
 C:Genetics:
 A:Map position: 4

Query Match 43.0%; Score 49; DB 2; Length 615;
 Best Local Similarity 71.4%; Pred. No. 50;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 KKIPEEEKKEEK 21
 |||:|||||
 Db 396 KKIPEEEKKEEK 409

RESULT 14
 S74279
 Hypothetical protein YCL061c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein YCL060c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 29-Oct-1999
 C:Accession: S74279; S19392; S19391; S29373; S21560
 R:Voet, M.; Volckaert, G.
 Submitted to the Protein Sequence Database, September 1996
 A:Reference number: S74277
 A:Accession: S74279

A;Molecule type: DNA
 A;Residues: 1-853 <VOE>
 A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42405.1; PID:e309033; PID:g190711
 A;Note: this is a revision to the sequence from reference S19391
 R;Rasmussen, S.W.; von Wettstein, D.
 submitted to the Protein Sequence Database, March 1992
 A;Reference number: S19391
 A;Accession: S19392
 A;Molecule type: DNA
 A;Residues: 'MKLPIMNHLVVMRTIVNPIVRKMKML',346-619,'LP',621,'P',623-638,'LRTIALM' <RAS>
 A;Cross-references: EMBL:X59720; MIPS:YCL061C
 A;Note: this sequence has been revised in reference S74279
 A;Note: this was assumed to be the complete sequence of protein YCL061C
 A;Accession: S19391
 A;Molecule type: DNA
 A;Residues: 1-314,'IEW' <RAW>
 A;Cross-references: EMBL:X59720
 A;Note: this sequence has been revised in reference S74279
 A;Note: this was assumed to be protein YCL060C
 R;Kern, L.
 Nucleic Acids Res. 18, 5279, 1990
 A;Title: The URK1 gene of *Saccharomyces cerevisiae* encoding uridine kinase.
 A;Reference number: S29373; MUID:90384830; PMID:2169608
 A;Accession: S29373
 A;Molecule type: DNA
 A;Residues: 417-504,'V',506-564,566-596 <KER>
 A;Cross-references: EMBL:X53998; NID:g4771; PIDN:CAA37945.1; PID:g4772
 C;Genetics:
 A;Map position: 3L
 A;Note: YCL061C

Query Match 43.0%; Score 49; DB 2; Length 853;
 Best Local Similarity 56.2%; Pred. No. 72;
 Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 7 HKKIPPEKREKLEQ 22
 Db 277 NKRIROKREKREKLE 292

RESULT 15

C71956
 Probable endopeptidase Clp ATP-binding chain - *Helicobacter pylori* (strain J99)
 N;Alternate names: ATP-dependent Clp proteinase regulatory chain
 N;Contains: adenosinetriphosphatase (EC 3.6.1.3)
 C;Species: *Helicobacter pylori*
 A;Variety: strain J99
 C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 19-Jan-2001
 C;Accession: C71956
 R;Alm, R.A.; Ling, L.S.L.; Mol, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A;Reference number: A71800; MUID:99120557; PMID:9923682
 A;Accession: C71956
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-856 <ARN>
 A;Cross-references: GB:AE001462; GB:AE001439; NID:g4154760; PIDN:AAD05825.1; PID:g415476
 A;Experimental source: strain J99
 C;Genetics:
 A;Gene: clpB
 C;Function:
 A;Description: allows clp to hydrolyze polypeptides and proteins, probably by a chaperon
 e activity; ATP hydrolysis is required for Clp hydrolysis of proteins but not of smaller
 C;Superfamily: endopeptidase Clp ATP-binding chain
 C;Keywords: ATP; hydrolase; molecular chaperone; nucleotide binding; P-loop
 F;204-211/Region: nucleotide-binding motif A (P-loop)
 F;271-276/Region: nucleotide-binding motif B
 F;605-612/Region: nucleotide-binding motif A (P-loop)
 F;673-678/Region: nucleotide-binding motif B
 F;210/Binding site: ATP (Lys) #status predicted
 F;611/Binding site: ATP (Lys) #status predicted

Query Match 43.0%; Score 49; DB 2; Length 856;
 Best Local Similarity 53.3%; Pred. No. 72;
 Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
 Oy 7 HKKIPPEKREKLE 21
 Db 500 YSKIPPEKREKLEQ 514

Search completed: July 16, 2003, 07:56:52
 Job time: 3.1294 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:42 ; Search time 0.572433 Seconds
(without alignments)
1594.036 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114
Sequence: 1 KLGFPAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	100.0	1167	1	ITAG_HUMAN
2	51	44.7	128	1	YQ40_BACAN
3	50	44.7	856	1	CLPB_HELPY
4	51	43.9	357	1	Y234_LISMO
5	50	43.9	505	1	SPKD_SYNY3
6	50	43.9	1034	1	ITAV_CHICK
7	50	43.9	1044	1	ITAV_MOUSE
8	50	43.9	1048	1	ITAV_HUMAN
9	49.5	43.4	959	1	VP2_BT3V
10	49	43.0	357	1	Y266_LISIN
11	49	43.0	853	1	YCG1_YEAST
12	49	43.0	856	1	CLPB_HELPY
13	48.5	42.5	407	1	M21_STRPY
14	48	42.1	573	1	ERGI_RAT
15	48	42.1	1526	1	TP2A_RAT
16	47	41.2	112	1	FLIN_AOUAE
17	47	41.2	199	1	HMGA_MOUSE
18	47	41.2	692	1	YKOC_YEAST
19	46	40.4	308	1	AMIP_STRPN
20	46	40.4	450	1	YKHS_YEAST
21	46	40.4	1407	1	TRHY_RABIT
22	45	39.5	101	1	RL12_METTL
23	45	39.5	147	1	RL9_THETN
24	45	39.5	333	1	AOX2_SOYBN
25	45	39.5	339	1	RLAO_PYRFU
26	45	39.5	527	1	ZIM2_HUMAN
27	45	39.5	1022	1	TPS3_PAROL
28	45	39.5	1022	1	TPS3_YEAST
29	45	39.5	1085	1	YAPA_SCHPO
30	44	38.6	72	1	DMS4_HUMAN
31	44	38.6	93	1	S108_HUMAN
32	44	38.6	184	1	YMKC_BACSU
33	44	38.6	349	1	ARSA_METUA

34	44	38.6	350	1	Y170_METUA	Q57634 methanococ
35	44	38.6	450	1	FEM2_RAT	Q9awr7 rattus norv
36	44	38.6	475	1	TPS1_PICAN	Q94213 picchia angu
37	44	38.6	477	1	YF45_AOUAE	Q67499 aquifex aeo
38	44	38.6	478	1	TPS1_CANAL	Q92410 candida alb
39	44	38.6	566	1	PEP_YEAST	P11491 saccharomyc
40	44	38.6	568	1	DISC_DROME	P23792 drosophila
41	44	38.6	1033	1	ITAB_MOUSE	Q9quuo mus musculi
42	44	38.6	1121	1	MYT1_HUMAN	Q01538 homo sapien
43	43.5	38.2	1863	1	BRCL_HUMAN	P38398 homo sapien
44	43	37.7	100	1	RR14_LOTUA	Q9bbct lotus japon
45	43	37.7	100	1	RR14_TOBAC	P06371 nicotiana t

ALIGNMENTS

RESULT 1
ITAG_HUMAN STANDARD; PRT; 1167 AA.
AC Q75578; Q9UH28;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-10 precursor.
GN ITGA10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular chondrocytes;
RX MEDLINE=98352078; PubMed=9685391;
RA Camper L., Hellman U., Lundgren-Akerlund E.;
RT "Isolation, cloning, and sequence analysis of the integrin subunit
RT alpha10, a beta1-associated collagen binding integrin expressed on
RT chondrocytes";
RL J. Biol. Chem. 273:20383-20389(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart, and Endothelial cells;
RX MEDLINE=20169197; PubMed=10702680;
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
RA Wang S.-X., Langley R., Kristensen G.W.;
RT "The integrin alpha10 subunit: expression pattern, partial gene
RT structure, and chromosomal localization";
RL Cytogenet. Cell Genet. 87:238-244(1999).
CC -1- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.
CC -1- DOMAIN: THE INTEGRIN (INTEGRIN) IS A VMFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 VMFA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF074015; AACF1952.1; -;
CC DR EMBL: AF112345; AAF21944.1; -;
CC EMBL: AF172723; AAF61638.1; -;
CC HSSP: P17301; IAOX.
CC Genew: HGNC:6135; ITGA10.

```

DR MIM: 604042; -
DR InterPro: IPR000413; Integrin_alpha.
DR InterPro: IPR002035; VWF_A.
DR Pfam: PF00092; vwa; 1.
DR Pfam: PF01839; FG-GAP; 5.
DR PRINTS: PRO1185; INTEGRIN.
DR PRINTS: PRO0453; VWFADOMAIN.
DR SMART: SM00191; Int. alpha; 5.
DR SMART: SM00377; VWA; 1.
DR PROSITE: PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE: PS50234; VWF_A; 1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium.
FT SIGNAL 1 22
FT CHAIN 23 1167
FT DOMAIN 23 1122
FT TRANSLEM 1123 1145
FT DOMAIN 1146 1167
FT REPEAT 38 97
FT REPEAT 97 97
FT REPEAT 97 97
FT REPEAT 167 350
FT REPEAT 365 427
FT REPEAT 428 482
FT REPEAT 483 545
FT REPEAT 546 605
FT REPEAT 608 660
FT REPEAT 1134 1140
FT CA_BIND 494 502
FT CA_BIND 558 566
FT CA_BIND 620 628
FT DISULFID 76 86
FT DISULFID 666 675
FT DISULFID 681 736
FT DISULFID 789 795
FT CARBOHYD 98 98
FT CARBOHYD 234 234
FT CARBOHYD 336 336
FT CARBOHYD 364 364
FT CARBOHYD 733 733
FT CARBOHYD 763 763
FT CARBOHYD 839 839
FT CARBOHYD 921 921
FT CARBOHYD 1011 1011
FT CARBOHYD 1018 1018
FT CARBOHYD 1039 1039
FT CONFLICT 844 844
FT CONFLICT 909 909
FT CONFLICT 926 926
SQ SEQUENCE 1167 AA; 127573 MW; AE7D3A1C25C1AEAD CRC64;

Query Match 100.0%; Score 114; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 4.6e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPEEKREKLEQ 22
Db 1146 KLGFFAHKKIPEEKREKLEQ 1167

```

```

RP SEQUENCE FROM N.A.
RA Okinka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kruma S., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.,
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF18935; AAP13645.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 128 AA; 14687 MW; D54C69BAD5A0AFA CRC64;

Query Match 44.7%; Score 51; DB 1; Length 128;
Best Local Similarity 50.0%; Pred. No. 2.2;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LGFFAHKKIPEEKREK 19
Db 9 MGFFGNKGRKPKSEKDEK 26

RESULT 3
CLPB_HELPY STANDARD; PRT; 856 AA.
ID CLPB_HELPY
AC P71404;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE C1PB protein.
OS CLPB OR HP0264.
GN Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC 11637;
RA Allan E.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RA MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khairat H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RA pylori."
RL Nature 388:539-547(1997).
CC -1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR
CC -1- ATP-DEPENDENT PROTEASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLPB/CLPB FAMILY:
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB046600; BAB17036.1; -
DR EMBL; D64005; BAA10726.1; -
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR003646; SH3_bac.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00287; SH3D; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein; transferase; Serine/threonine-protein kinase;
KW ATP-binding; Complete proteome.
FT DOMAIN 9 271 PROTEIN KINASE.
FT NP_BIND 15 23 ATP (BY SIMILARITY).
FT BINDING 40 40 ATP (BY SIMILARITY).
FT ACT_SITE 136 136 BY SIMILARITY.
SQ SEQUENCE 505 AA; 55213 MW; C4F12A1886C4D51C CRC64;

Query Match 43.9%; Score 50; DB 1; Length 505;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 6 AHKKIPKREKREKLEQ 22
Db 376 AEOKIAENKREKRELEQ 392

RESULT 6
ITAV_CHICK STANDARD; PRT; 1034 AA.
ID ITAV_CHICK
AC P26008;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-V precursor (Vltroectin receptor alpha subunit).
GN ITGAV.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
CX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=91104936; PubMed=1703004;
RA Borey B.; Reichardt L.F.;
RT "Chick integrin alpha V subunit molecular analysis reveals high
RT conservation of structural domains and association with multiple beta
RT subunits in embryo fibroblasts.";
RL Biochemistry 29:10191-10198(1990).
CC -1- FUNCTION: THE ALPHA-V INTEGRINS ARE RECEPTORS FOR VITROECTIN,
CC CYTOACTIN, FIBRONECTIN, FIBRINOGEN, LAMININ, MATRIX
CC METALLOPROTEINASE-2, OSTEOPOINTIN, PROTHROMBIN, THROMBOSPONDIN AND
CC VON WILLEBRAND FACTOR. THEY RECOGNIZE THE SEQUENCE R-G-D IN A WIDE
CC ARRAY OF LIGANDS.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,
CC BETA-5, BETA-6 OR BETA-8.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M60517; AAA49138.1; -
DR PIR; A36108; A36108.
DR HSSP; P06756; 1JV2.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRIN.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium.
FT SIGNAL 1 19
FT CHAIN 20 1034
FT CHAIN 20 875
FT CHAIN 20 978
FT DOMAIN 20 978
FT TRANSMEM 979 1002
FT DOMAIN 1003 1034
FT REPEAT 35 96
FT REPEAT 101 166
FT REPEAT 167 224
FT REPEAT 237 290
FT REPEAT 291 350
FT REPEAT 355 415
FT REPEAT 420 473
FT CA_BIND 248 256
FT CA_BIND 302 310
FT CA_BIND 367 375
FT CA_BIND 431 439
FT SITE 1005 1009
FT DISULFID 77 85
FT DISULFID 126 146
FT DISULFID 160 173
FT DISULFID 479 488
FT DISULFID 494 551
FT DISULFID 612 618
FT DISULFID 684 697
FT DISULFID 838 890
FT DISULFID 895 900
FT CARBOHYD 62 62
FT CARBOHYD 278 278
FT CARBOHYD 284 284
FT CARBOHYD 540 540
FT CARBOHYD 601 601
FT CARBOHYD 690 690
FT CARBOHYD 821 821
FT CARBOHYD 837 837
FT CARBOHYD 860 860
FT CARBOHYD 931 931
FT CARBOHYD 951 951
FT CARBOHYD 959 959
FT CARBOHYD 966 966
SQ SEQUENCE 1034 AA; 114388 MW; D76B08BA692DC684 CRC64;

Query Match 43.9%; Score 50; DB 1; Length 1034;
Best Local Similarity 38.1%; Pred. No. 26;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 1 KLGFPAHKIPKREKREKLE 21
Db 1003 RMGFPAKVRPPQREOREBOLQ 1023

RESULT 7
ITAV_MOUSE STANDARD; PRT; 1044 AA.
ID ITAV_MOUSE
AC P43406;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

RP SEQUENCE OF 31-41.
 RX MEDLINE=89195223; PubMed=2467745;
 RA Chershe D.A., Smith J.W., Cooper H.M., Quaranta V.;
 RT "A novel vitronectin receptor integrin (alpha v beta x) is
 RT responsible for distinct adhesive properties of carcinoma cells.";
 RT Cell 57:59-69(1989).
 RA [5]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 31-987.
 RX MEDLINE=21482770; PubMed=11546839;
 RA Xiong J.P., Stehle T., Dieffenbach B., Zhang R., Dunker R., Scott D.L.,
 RA Joachimiak A., Goodman S.L., Arnaout M.A.;
 RT "Crystal structure of the extracellular segment of integrin alpha
 RT Vbeta3.";
 RT Science 294:339-345(2001).
 CC -!- FUNCTION: The alpha-V integrins are receptors for vitronectin,
 CC cytoactin, fibronectin, fibrinogen, laminin, matrix
 CC metalloproteinase-2, osteopontin, osteomodulin, prothrombin,
 CC thrombospondin and von Willebrand factor. They recognize the
 CC sequence R-G-D in a wide array of ligands.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
 CC DISULFIDE BOND. ALPHA-V ASSOCIATES WITH EITHER BETA-1, BETA-3,
 CC BETA-5, BETA-6 OR BETA-8 SUBUNIT.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -!- DATABASE: NAME=PROV; NOTE=CD guide CDS1 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cds1.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M14648; AAA36808.1; -;
 DR EMBL; AF251841; AAG03000.1; JOINED.
 DR EMBL; AF251818; AAG03000.1; JOINED.
 DR EMBL; AF251819; AAG03000.1; JOINED.
 DR EMBL; AF251820; AAG03000.1; JOINED.
 DR EMBL; AF251821; AAG03000.1; JOINED.
 DR EMBL; AF251822; AAG03000.1; JOINED.
 DR EMBL; AF251823; AAG03000.1; JOINED.
 DR EMBL; AF251824; AAG03000.1; JOINED.
 DR EMBL; AF251825; AAG03000.1; JOINED.
 DR EMBL; AF251826; AAG03000.1; JOINED.
 DR EMBL; AF251827; AAG03000.1; JOINED.
 DR EMBL; AF251828; AAG03000.1; JOINED.
 DR EMBL; AF251829; AAG03000.1; JOINED.
 DR EMBL; AF251830; AAG03000.1; JOINED.
 DR EMBL; AF251831; AAG03000.1; JOINED.
 DR EMBL; AF251832; AAG03000.1; JOINED.
 DR EMBL; AF251833; AAG03000.1; JOINED.
 DR EMBL; AF251834; AAG03000.1; JOINED.
 DR EMBL; AF251835; AAG03000.1; JOINED.
 DR EMBL; AF251836; AAG03000.1; JOINED.
 DR EMBL; AF251837; AAG03000.1; JOINED.
 DR EMBL; AF251838; AAG03000.1; JOINED.
 DR EMBL; AF251839; AAG03000.1; JOINED.
 DR EMBL; AF251840; AAG03000.1; JOINED.
 DR PIR; A27421; A27421.
 DR PIR; A26482; A26482.
 DR PIR; A32287; A32287.
 DR PIR; B32287; B32287.
 DR PDB; 1UV2; 17-OCT-01.
 DR Genew; HGNC:6150; ITGAV.
 DR MIM; 193210; -;
 DR InterPro; IPR000413; Integrin_alpha.
 DR Pfam; PF00357; Integrin_A_1.
 DR Pfam; PF01839; FG-GAP_5.
 DR PRINTS; PRO1185; INTEGRINA.
 DR SMART; SMO0191; int_alpha; 5.

DR PROSITE; PS00242; INTEGRIN ALPHA; 1.
 KW Integrin; Cell adhesion; Glycoprotein; Transmembrane; Signal;
 KM Repeat; Calcium; 3D-structure.
 FT SIGNAL 1 30
 FT CHAIN 31 1048 INTEGRIN ALPHA-V.
 FT CHAIN 31 889 INTEGRIN ALPHA-V HEAVY CHAIN.
 FT CHAIN 891 1048 INTEGRIN ALPHA-V LIGHT CHAIN.
 FT DOMAIN 31 992 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 993 1016 POTENTIAL.
 FT DOMAIN 1017 1048 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 46 108 FG-GAP 1.
 FT REPEAT 113 178 FG-GAP 2.
 FT REPEAT 179 236 FG-GAP 3.
 FT REPEAT 249 302 FG-GAP 4.
 FT REPEAT 303 362 FG-GAP 5.
 FT REPEAT 367 427 FG-GAP 6.
 FT REPEAT 432 483 FG-GAP 7.
 FT CA_BIND 260 268 POTENTIAL.
 FT CA_BIND 314 322 POTENTIAL.
 FT CA_BIND 379 387 POTENTIAL.
 FT CA_BIND 443 451 POTENTIAL.
 FT SITE 1019 1023 GPEPR MOTIF.
 FT DISULFID 89 97 BY SIMILARITY.
 FT DISULFID 138 158 BY SIMILARITY.
 FT DISULFID 172 185 BY SIMILARITY.
 FT DISULFID 491 502 BY SIMILARITY.
 FT DISULFID 508 565 BY SIMILARITY.
 FT DISULFID 626 632 BY SIMILARITY.
 FT DISULFID 698 711 BY SIMILARITY.
 FT DISULFID 852 904 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 909 914 BY SIMILARITY.
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 704 704 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 425 425 W -> R (IN REF. 2).
 FT CONFLICT 783 783 I -> V (IN REF. 2).
 FT CONFLICT 1039 1039 H -> R (IN REF. 2).
 SQ SEQUENCE 1048 AA; 116051 MW; 3648B0B350D6D07 CRC64;
 Query Match 43.9%; Score 50; DB 1; Length 1048;
 Best Local Similarity 38.1%; Pred. No. 26;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;
 QY 1 KLGFPAAHKIPDEKREKLE 21
 Db 1017 RMGFFKVRPPQEEQERQLQ 1037
 RESULT 9
 ID_VP2_BTIV3V STANDARD; PRT; 959 AA.
 AC 006998;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Outer capsid protein VP2.
 GN S2.
 OS Bluecough virus (serotype 3 / isolate South Africa-vaccine).
 OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
 OC NCBI_TaxID=36424;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91021485; PubMed=2171239;

RA Gould A.R., Prichard L.I.;
RT "Relationships amongst bluetongue viruses revealed by comparisons of
RT capsid and outer coat protein nucleotide sequences.";
RL Virus Res. 17:31-52(1990).
CC -!- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VPS)
CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55801; CAA39323.1; -;
DR PIR; B60017; B60017;
DR InterPro; IPR001742; Orbi_VP2.
DR Pfam; PF00898; Orbi_VP2; 1.
DR ProDom; PD002938; Orbi_VP2; 1.
KW Coat protein.
SQ SEQUENCE 959 AA; 112163 MW; 41B230E31803588F CRC64;

Query Match 43.4%; Score 49.5; DB 1; Length 959;
Best Local Similarity 57.1%; Pred. No. 28;
Matches 12; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 2 LGFPAHKKIPBEKREKLEQ 22
Db 791 LNFF-----PSYERREKLEK 806
||| |||||
Y266 LISIN STANDARD; PRT; 357 AA.
ID Y266 LISIN STANDARD; PRT; 357 AA.
AC Q92F41;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Lin0266.
GN Lin0266.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Francaul L., Buchrieser C., Ruenick C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Checouani F., Couve E., de Darvar A., Denoux F.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Eutlier K.-D., Feihl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitounan A., Mata Vicente J., Ng E., Medjari H.,
RA Nordstedt G., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- SIMILARITY: BELONGS TO THE YCF81 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; A1596164; CAC95499.1; -;
DR Listlist; LIN00266; -;
DR InterPro; IPR002792; TRAM.
DR Pfam; PF01938; TRAM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 357 AA; 39433 MW; 23A6C248979167F4 CRC64;

Query Match 43.0%; Score 49; DB 1; Length 357;
Best Local Similarity 71.4%; Pred. No. 12;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 AHKKIPBEKREK 19
Db 146 AKKTPBEKREK 159
||| |||||
YCG1 YEAST STANDARD; PRT; 853 AA.
ID YCG1 YEAST STANDARD; PRT; 853 AA.
AC P25588; P25589; P27513; P87003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 97.9 kDa protein in CHAI-KRI intergenic region.
GN YCL061C OR YCL61C/YCL60C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN RN
RP SEQUENCE FROM N.A.
RA Rasmussen S.W., von Wettstein D.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN RN
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN RN
RP SEQUENCE OF 384-596 FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE=90384830; PubMed=2169608;
RA Kern L.;
RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
RL Nucleic Acids Res. 18:5279-5279(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42405.1; -;
DR EMBL; X53998; -; NOT_ANNOTATED_CDS.
DR PIR; S19391; S19391.
DR PIR; S19392; S19392.
DR PIR; S21360; S21360.
DR PIR; S29373; S29373.
DR SGD; S0000566; YCL061C.
KW Hypothetical protein.
FT CONFLICT 505 505 L -> V (IN REF. 3).
FT CONFLICT 567 567 MISSING (IN REF. 3).
SQ SEQUENCE 853 AA; 97946 MW; 16E09FCC0B246D1 CRC64;

Query Match 43.0%; Score 49; DB 1; Length 853;
Best Local Similarity 56.2%; Pred. No. 29;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 7 HKKIPBEKREKLEQ 22
Db 277 NKRIROKREKREKLE 292
||| |||||

ID	CLPB_HELPF	STANDARD:	PRT:	856 AA.
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	CLPB protein.			
GN	CLPB OR JHP0249.			
OS	Helicobacter pylori J99 (Campylobacter pylori J99)			
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
OC	Helicobacter.			
OX	NCBI_TaxID=85963;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99120557; PubMed=9923682;			
RA	Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,			
RA	Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,			
RA	Thumming P.J., Casuso A., Uria-Nickelsen M., Mills D.M., Ives C.,			
RA	Gibson R.J., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,			
RT	Trust T.J.;			
RT	"Genomic sequence comparison of two unrelated isolates of the human			
RL	gastric pathogen Helicobacter pylori.";			
RL	Nature 397:176-180 (1999).			
CC	-1- SUBUNIT: THOUGHT TO BE AN ATPASE SUBUNIT OF AN INTRACELLULAR			
CC	ATP-DEPENDENT PROTEASE (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-slb.ch/announce/			
CC	or send an email to license@isb-slb.ch).			
CC	-----			
DR	EMBL; AE01462; AAD05825.1; -			
DR	InterPro; IPR003593; AAA ATPase.			
DR	InterPro; IPR003959; AAA ATPase centr.			
DR	InterPro; IPR001270; Chaprinin_clpb/B.			
DR	InterPro; IPR004176; CLP_N.			
DR	Pfam; PF00004; AAA; 2.			
DR	Pfam; PF02861; CLP_N; 2.			
DR	PRINTS; PR00300; CLPBPROTEASEA.			
DR	ProDom; PD000739; GSP1I_E; 1.			
DR	SMART; SM00382; AAA; 2.			
DR	PROSITE; PS00870; CLPBAB_1; 1.			
DR	PROSITE; PS00871; CLPBAB_2; 1.			
KW	Chaperone; ATP-binding; Repeat; Complete proteome.			
FT	DOMAIN 159 406			
FT	DOMAIN 531 722			
FT	NP_BIND 204 211			
FT	NP_BIND 605 612			
FT	SEQUENCE 856 AA; 96730 MW; 086658F6946588A CRC64;			
Qy	7 HKKIPPEKREKLE 21			
Db	500 YSKIPEKEKEKELEQ 514			
Query Match	43.0%; Score 49; DB 1; Length 856;			
Best Local Similarity	53.3%; Pred. No. 29;			
Matches	8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;			
RESULT 13				
M21_STRPY	STANDARD;	PRT;	407 AA.	
AC	P50468;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	M protein, serotype 2.1 precursor.			

```

NN EMBL2.1
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=72/44/RB4;
RX MEDLINE=92104662; PubMed=1370269;
RA Bessen D.E., Fischetti V.A.;
RT "Identification of a unique immunoglobulin A-binding protein.";
RL Infect. Immun. 60:124-135(1992).
RN [2]
RP SEQUENCE OF 138-305 FROM N.A.
RC STRAIN=72/44/RB4;
RX MEDLINE=91079780; PubMed=2258705;
RA Bessen D.E., Fischetti V.A.;
RT "Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules.";
RL J. Exp. Med. 172:1757-1764(1990).
CC -1- FUNCTION. THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO PHAGOCYTOSIS.
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (potential).
CC -1- SIMILARITY: TO OTHER M PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb.ch/announce or send an email to license@isb.ch).
CC -----
DR EMBL; X61276; CAA43581.1; -;
DR EMBL; X56398; CAA39808.1; -;
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003345; M_repeat.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02370; M_9.
DR PRINTS; PR00015; GPOSANCHOR.
DR TIGRFAMs; TIGR01167; LPTXG_anchor; 1.
DR TIGRFAMs; TIGR01168; YSIRK_signal; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
KW Antigen; Coiled coil; Signal.
FT SIGNAL 1 41
FT CHAIN 42 377
FT PROPEP 378 407
FT DOMAIN 81 94
FT REPEAT 81 87
FT REPEAT 88 94
FT DOMAIN 158 285
FT REPEAT 158 180
FT REPEAT 193 215
FT REPEAT 228 250
FT REPEAT 263 285
FT DOMAIN 116 157
FT DOMAIN 348 373
FT SITE 374 378
FT MOD_RES 377 377
SQ SEQUENCE 407 AA; 46466 MW; 33CA053B7DB3C1EA CRC64;
Query March 42.5%; Score 48.5; DB 1; Length 407;
Best local Similarity 58.8%; Pred. No. 16;
Matches 10; Conservative 4; Mismatches 2; Indels 1; Gaps 1.

```

QY 7 HKKPEEKRE-EKLEQ 22
 DB 85 HKKVEEHKKDHEKLEK 101

RESULT 14

ERG1_RAT STANDARD; PRT; 573 AA.
 AC P52020;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Squalene monooxygenase (EC 1.14.99.7) (Squalene epoxidase) (SE).
 GN SOLE OR ERG1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=95113819; PubMed=7814369;
 RA Sakakibara J, Matanabe R, Kanai Y, Ono T.;
 RT "Molecular cloning and expression of rat squalene epoxidase.";
 RL J. Biol. Chem. 270:17-20(1995).
 CC -1- FUNCTION: CATALYZES THE FIRST OXYGENATION STEP IN STEROL
 BIOSYNTHESIS AND IS SUGGESTED TO BE ONE OF THE RATE-LIMITING
 ENZYMES IN THIS PATHWAY.
 CC -1- CATALYTIC ACTIVITY: Squalene + AH(2) + O(2) = (S)-squalene-2,3-
 epoxide + A + H(2)O.
 CC -1- COFACTOR: FAD.
 CC -1- SUBUNIT: MAY FORM A COMPLEX WITH SQUALENE SYNTHASE.
 CC -1- SUBCELLULAR LOCATION: Microsomal.
 CC -1- SIMILARITY: BELONGS TO THE SQUALENE MONOOXYGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D37920; BA07141.1; -
 DR InterPro: IPR000733; Flav monooxygenase.
 DR InterPro: IPR000205; NAD binding.
 DR Pfam; PF01360; Monooxygenase; 1.
 DR Oxidoreductase; Flavoprotein; FAD; Transmembrane.
 FT TRANSMEM 20 40 POTENTIAL.
 FT TRANSMEM 60 80 POTENTIAL.
 FT TRANSMEM 118 138 POTENTIAL.
 FT TRANSMEM 545 565 POTENTIAL.
 FT NP_BIND 125 152 FAD (ADP PART) (POTENTIAL).
 FT NP_BIND 152 152 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 573 AA; 64024 MW; 55222C911E2B8777 CRC64;

Query Match 42.1%; Score 48; DB 1; Length 573;
 Best Local Similarity 52.9%; Pred. No. 27;
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LGFFAHKKIPEEKRE 18
 DB 75 IGFFWAKSPSEKKEQ 91

RESULT 15

TP2A_RAT STANDARD; PRT; 1526 AA.
 AC P41516;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA topoisomerase II, alpha isozyme (EC 5.99.1.3).
 GN TOP2A OR TOP2 OR TOP2-2.

OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;
 RX MEDLINE=93290677; PubMed=8390253;
 RA Park S.H., Yoon J.H., Kwon Y.D., Park S.D.;
 RT "Nucleotide sequence analysis of the cDNA for rat DNA topoisomerase
 II.";
 RL Biochem. Biophys. Res. Commun. 193:787-793(1993).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
 of double-stranded DNA.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; Z46372; CA66496.1; -
 DR EMBL; Z19552; CA79611.1; -
 DR HSSP; Z28676; -, NOT_ANNOTATED_CDS.
 DR HSSP; P06786; IBGW.
 DR InterPro: IPR003594; ATPbind ATPase.
 DR InterPro: IPR003957; CBFA NFYB topis.
 DR InterPro: IPR001241; DNA_TopoisoII.
 DR InterPro: IPR002205; DNA_topoisoIV.
 DR Pfam; PF00204; DNA_gyraseB; 1.
 DR Pfam; PF00521; DNA_topoisoIV; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR PRINTS; PR00615; CCAATSUBUNTA.
 DR PRINTS; PR00418; TP12PFAMLY.
 DR ProDom; PD000616; DNA_topoisoII; 1.
 DR ProDom; PD000742; DNA_topoisoII; 1.
 DR SMART; SM00433; TOP2c; 1.
 DR SMART; SM00434; TOP4c; 1.
 DR PROSITE; PS00177; TOPOISOMERASE_II; 1.
 KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 FT NP_BIND 153 164 ATP (POTENTIAL).
 FT ACT_SITE 803 803 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1526 AA; 173220 MW; A1961ABDB1B050F CRC64;

Query Match 42.1%; Score 48; DB 1; Length 1526;
 Best Local Similarity 56.2%; Pred. No. 72;
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKPEEKREKLE 21
 DB 1091 AQQKVPDEENENEE 1106

Search completed: July 16, 2003, 07:52:53
 Job time : 2.57243 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:48:02 ; Search time 2.25879 Seconds
(without alignments)
2006.842 Million cell updates/sec

```
Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFPAHKCIPEEKREKLEQ 22
```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

```
Database : SPREMBL_21:*
1 : sp_archaea:*
2 : sp_bacteria:*
3 : sp_fungi:*
4 : sp_human:*
5 : sp_invertebrate:*
6 : sp_mammal:*
7 : sp_mhc:*
8 : sp_organelle:*
9 : sp_phage:*
10 : sp_plant:*
11 : sp_rodent:*
12 : sp_virus:*
13 : sp_vertebrate:*
14 : sp_unclassified
15 : sp_virus:*
16 : sp_bacteriophage:*
17 : sp_archaeo:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	46.5	214	16	08ZV0	08rjv0 salmonella
2	53	46.5	214	16	08Z0T9	08z0t9 salmonella
3	53	46.5	1033	13	042598	042598 xenopus la
4	52	45.6	1021	5	09GSR4	09gsr4 podocoryne
5	51	44.7	180	17	08RTJ6	08rtj6 pyrococcus
6	51	44.7	672	17	028247	028247 archaeosol
7	51	44.7	1045	13	091292	091292 pleurodele
8	51	44.7	1957	11	091XC9	091xc9 mus muscul
9	50	43.9	403	5	09X1F8	09x1f8 trypanosoma
10	50	43.9	1007	6	09GK48	09gk48 bos taurus
11	50	43.9	1047	6	09MZD6	09mzd6 bos taurus
12	49.5	43.4	236	9	064016	064016 bacteriophan
13	49.5	43.4	236	16	032005	032005 bacillus s
14	49.5	43.4	635	5	0961R8	0961r8 homo sapien
15	49.5	43.4	1372	4	09HAN1	09han1 homo sapien
16	49.5	43.4	1388	4	09HAW2	09haw2 homo sapien

17	49.5	43.4	2187	4	Q9H137	Q9h197 homo sapien
18	49.5	43.4	2254	4	Q9HC10	Q9h190 homo sapien
19	49.5	43.0	615	10	Q9SMQ3	Q9smq3 arabidopsi
20	49	43.0	2527	5	Q5W63	Q5w683 plasmodi
21	48.5	42.5	211	2	Q554B2	Q554b2 streptococc
22	48	42.1	401	5	Q9U9N9	Q9u9n9 trypanosoma
23	48	42.1	421	17	Q97VL7	Q97vl7 sulfobolus
24	48	42.1	936	12	Q8RV7	Q8rv7 chimpanzee
25	48	42.1	958	10	Q9AVP6	Q9avp6 vicia faba
26	48	42.1	1034	6	Q9TUN4	Q9tun4 oryctolagus
27	48	42.1	1070	12	Q8QJ26	Q8qj26 sulfobolus
28	48	42.1	1664	5	Q56782	Q56782 dirosophila
29	48	42.1	1654	5	Q9VX75	Q9vx75 dirosophila
30	48	42.1	3487	2	Q9FDU0	Q9fd0 microcystis
31	48	42.1	3487	2	Q9NB3	Q9nbd3 microcystis
32	47.5	41.7	2343	6	Q18806	Q18806 canis famli
33	47	41.2	102	17	Q9UZx3	Q9uzx3 pyrococcus
34	47	41.2	208	5	Q9GY4	Q9gy4 caenorhabdi
35	47	41.2	306	10	Q9SRV7	Q9sr7 arabidopsi
36	47	41.2	309	4	Q9H4K1	Q9h4k1 homo sapien
37	47	41.2	512	16	Q8UDL7	Q8udl7 agrobacter
38	47	41.2	702	4	Q9NVC7	Q9nvc7 homo sapien
39	47	41.2	702	4	Q9H8Y1	Q9h8y1 homo sapien
40	47	41.2	861	10	Q80738	Q80738 arabidopsi
41	46.5	40.8	170	17	Q96Z70	Q96z70 sulfobolus
42	46.5	40.4	170	9	Q9ZXJ6	Q9zxj6 bacteriophus
43	46	40.4	175	9	Q9CMW9	Q9cm9 streptococc
44	46	40.4	278	10	Q9S186	Q9s186 arabidopsi
45	46	40.4	292	5	Q94824	Q94824 tetrahymena

ALIGNMENTS

RESULT 1

ID	Q8ZJV0	PRELIMINARY;	PRT;	214 AA.
----	--------	--------------	------	---------

DT 01-MAR-2002 (T-EMBLrel, 20, Created)
DT 01-MAR-2002 (T-EMBLrel, 20, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel, 20, Last annotation update)
DE Membrane protein, transcribed divergently from serb.
DE SMP OR SPM4577.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxId=602;
X

```

Query Match      46.5%  Score 53;  DB 16;  length 214;
Best Local Similarity 71.4%  Pred. No. 8.3;
Matches 10;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0

```

```
QY      6 AHKKIPEEEKREEK 19
      | | : | | | : | | |
Db     200 ASKPVPPEEEREK 213
```

RESULT 2

```
082079
ID 082079 PRELIMINARY; PRT: 214 AA.
AC 082079
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Putative membrane protein.
DE SMP OR STY4924.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jazayeri C.,
RA Krogsh A., Larsen T.S., Leather S., Moutie S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL677284; CAD03408.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 214 AA; 24279 MW; 2683D0DF9174BF80 CRC64;

Query Match 46.5%; Score 53; DB 16; Length 214;
Best Local Similarity 71.4%; Pred. No. 8.3;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREK 19
Db 200 ASKPEPEEREK 213

RESULT 3
ID 042598 PRELIMINARY; PRT: 1033 AA.
AC 042598;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE "Cloning of the Xenopus integrin alpha(v) subunit and analysis of its
DE distribution during early development."
OS Integrin alpha(v) subunit.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xeropodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=98211587; PubMed=9551862;
RA Joss T.O., Reintsch W.E., Brinker A., Klein C., Hausen P.;
RT "Cloning of the Xenopus integrin alpha(v) subunit and analysis of its
RT distribution during early development."
RL Int. J. Dev. Biol. 42:171-179(1998).
DR EMBL; U92006; AAB62090.1; -
DR HSP; P06756; IUV2.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
SQ SEQUENCE 1033 AA; 114025 MW; 35F73B0765C06BD6 CRC64;

Query Match 46.5%; Score 53; DB 13; Length 1033;
Best Local Similarity 47.6%; Pred. No. 37;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
```

```
QY 1 KLGFAFKKIPEEKREK 21
Db 1002 KLGFFKVRPPQETTEREQ 1022

RESULT 4
ID 09GSF4 PRELIMINARY; PRT: 1021 AA.
AC 09GSF4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE Integrin alpha chain.
DE INTRA.
OS Podocoryne carnea.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;
OC Hydrozoa; Podocoryne.
OX NCBI_TaxID=6096;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21375962; PubMed=11482899;
RA Reber-Mueller S., Studer R., Muller P., Yanze N., Schmid V.;
RT "Integrin and talin in the jellyfish Podocoryne carnea."
RL Cell Biol. Int. 25:753-769(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Reber-Mueller S., Studer R., Muller P., Yanze N., Schmid V.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF308651; AAG25933.1; -
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN 1.
SQ SEQUENCE 1021 AA; 110987 MW; 28A8E782FF616DB2 CRC64;

Query Match 45.6%; Score 52; DB 5; Length 1021;
Best Local Similarity 55.6%; Pred. No. 50;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KLGFAFKKIPEEKREK 18
Db 996 KKGFFKRXKXGDEDEEE 1013

RESULT 5
ID 08TZ06 PRELIMINARY; PRT: 180 AA.
AC 08TZ06;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE "The complete sequence of the Pyrococcus furiosus genome."
DE Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
OS Pyrococcus.
OS Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010292; AAB82119.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 180 AA; 21325 MW; 33BAE26104B200AE CRC64;

Query Match 44.7%; Score 51; DB 17; Length 180;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

QY 2 LGFFFAKKIPEEKRE 16
 Db 122 VGFPIHKKINEEBAK 136

RESULT 6

028247 PRELIMINARY; PRT; 672 AA.
 AC 028247;
 DT 01-JAN-1998 (TEMBUREL 05, Created)
 DT 01-JAN-1998 (TEMBUREL 05, Last sequence update)
 DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)
 DE Signal-transducing histidine kinase, putative.
 GN AF2032.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 OK NCBI_TaxID=2234;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; Pubmed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kellavage A.R., Graham D.E., Kyriades N.C.,
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Cotton M.D., Spriggs T., Arlisch P., Kaine B.P., Sykes S.M.,
 RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 350:364-370(1997).
 DR EMBL; AE000963; AAB89224.1; -.
 DR TIGR; AF2032; -.
 DR InterPro: IPR003661; His_kinA.
 DR InterPro: IPR004359; HIS_KIN_eig.
 DR InterPro: IPR001610; PAC.
 DR InterPro: IPR007000; PAS-aseoc-C.
 DR InterPro: IPR000014; PAS_domain.
 DR Pfam; PF00785; PAC; 3.
 DR Pfam; PF00989; PAS; 4.
 DR Pfam; PF00512; signal; 1.
 DR SMART; SM00388; HiskA; 1.
 DR SMART; SM00086; PAC; 3.
 DR SMART; SM00091; PAS; 4.
 DR TIGRFAMs; TIGR00229; sensory_box; 4.
 DR Hypochemical protein; Complete proteome.
 KW HYPOTHETICAL PROTEIN; SENSORY BOX; 4.
 SQ SEQUENCE 672 AA; 78620 MW; F95DC675BDB8F749 CRC64;

Query Match

Best Local Similarity 44.7%; Score 51; DB 17; Length 672;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPEEKREKLEQ 22
 Db 572 FAFKSWEDERRRELLKQ 589

RESULT 7

091292 PRELIMINARY; PRT; 1045 AA.
 AC 091292;
 DT 01-NOV-1996 (TEMBUREL 01, Created)
 DT 01-NOV-1996 (TEMBUREL 01, Last sequence update)
 DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)
 DE Integrin.
 OS Pleurodeles waltl (Tiberian ribbed newt).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae;

OC Pleurodeles.
 OK NCBI_TaxID=8319;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95377519; Pubmed=7649360;
 RA Alfandari D., Whitaker C.A., Desimone D.W., Darrivere T.,
 RT "Integrin alpha v subunit is expressed on mesodermal cell surfaces
 RT during amphibian gastrulation.";
 RL Dev. Biol. 170:249-261(1995).
 DR EMBL; X81108; CA57014.1; -.
 DR HSSP; P06756; IJVT.
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00191; Int_alpha; 5.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
 KW Integrin.
 SQ SEQUENCE 1045 AA; 115421 MW; 4DB34B766B0C648B CRC64;

Query Match 44.7%; Score 51; DB 13; Length 1045;
 Best Local Similarity 42.9%; Pred. No. 71;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 KLGFPAKKIPEEKREKLE 21
 Db 1014 KFGFFKRVBPQEGQERQLO 1034

RESULT 8

091Y9C PRELIMINARY; PRT; 1957 AA.
 AC 091Y9C;
 DT 01-DEC-2001 (TEMBUREL 19, Created)
 DT 01-DEC-2001 (TEMBUREL 19, Last sequence update)
 DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)
 DE 53Bp1 protein.
 GN TRP53BP1 OR 53BP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB C;
 RA Adachi Y.; (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RL Submitted [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=BALB C;
 RA Jullien D.;
 RT "kinetochore localisation of the DNA damage response component 53BP1
 RT during mitosis.";
 RL J. Cell Sci. 0:0-0(0).
 DR EMBL; AJ414734; CAC94013.1; -.
 DR MGD; MG11351320; TRP53bp1.
 DR InterPro: IPR001357; BRCt.
 DR InterPro: IPR002114; HPr_Serp_site.
 DR Pfam; PF00533; BRCt; 2.
 DR PROSITE; PS00172; BRCt; 2.
 DR PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
 SQ SEQUENCE 1957 AA; 211438 MW; 9199C7F4B627A8F CRC64;

Query Match 44.7%; Score 51; DB 11; Length 1957;
 Best Local Similarity 69.2%; Pred. No. 1.3e+02;
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 10 IPPEEKREKLEQ 22
 Db 129 LPPEEKREBELR 141

RESULT 9

09xyf8 PRELIMINARY; PRT; 403 AA.
 ID 09xyf8
 AC 09xyf8
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Calreticulin.
 GN CLR.
 OS Trypanosoma cruzi.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5693;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TULAHUEN 2;
 RX MEDLINE=99250150; PubMed=10233151;
 RA Labriola C., Gazzulo J.J., Parodi A.J.;
 RT "Trypanosoma cruzi calreticulin is a lectin that binds
 RT monoglucosylated oligosaccharides but not protein moieties of
 RT glycoproteins."
 RL Mol. Biol. Cell. 10:1381-1394(1999).
 DR EMBL; AF107115; AAD22175.1; -
 DR InterPro; IPR001580; Calreticulin.
 DR Pfam; PF00262; Calreticulin; 1.
 DR PRINTS; PR00626; CALRETICULIN.
 DR ProDom; PD001866; Calreticulin; 1.
 DR PROSITE; PS00803; CALRETICULIN_1; 1.
 DR PROSITE; PS00804; CALRETICULIN_2; UNKNOWN_1.
 DR PROSITE; PS00805; CALRETICULIN_REPEAT; 1.
 SQ SEQUENCE 403 AA; 46758 MW; AYBD560291AD071F CRC64;

Query Match 43.9%; Score 50; DB 5; Length 403;
 Best Local Similarity 66.7%; Pred. No. 40;
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 8 KTIPEEKREKLEHQ 22
 Db 380 KENPEKDEBELE 394

RESULT 10
 09GK48
 ID 09GK48 PRELIMINARY; PRT; 1007 AA.
 AC 09GK48
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Integrin alpha-V subunit (Fragment).
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY GLAND;
 RA Andersen M.H., Rasmussen J.T., Berglund L., Petersen T.E.;
 RT "Bovine alpha-V integrin subunit (fragment).";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF317199; AAG38595.1; -
 DR HSSP; P06756; IJV2.
 DR InterPro; IPR000413; Integrin_alpha.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00191; Int_alpha; 5.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 FT NON TER 1
 SQ SEQUENCE 1007 AA; 111460 MW; 51AFBC6DFCC36761 CRC64;

Query Match 43.9%; Score 50; DB 6; Length 1007;
 Best Local Similarity 38.1%; Pred. No. 95;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAAKKIPEEKREKLE 21
 Db 976 RMGFFKVRPQEQEREDQ 996

RESULT 11
 09MZD6 PRELIMINARY; PRT; 1047 AA.
 ID 09MZD6
 AC 09MZD6
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Integrin alpha V subunit precursor.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RX MEDLINE=20366286; PubMed=10906183;
 RA Neff S., Mason P.W., Baxt B.;
 RT "High-efficiency utilization of the bovine integrin alpha(v)beta(3) as
 RT a receptor for foot-and-mouth disease virus is dependent on the bovine
 RT beta(3) subunit."
 RL J. Virol. 74:7298-7306(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG;
 RA Neff S., Mason P.W., Baxt B.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF239958; AAF44691.2; -
 DR HSSP; P06755; IJV2.
 DR InterPro; IPR000413; Integrin_alpha.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR SMART; SM00191; Int_alpha; 4.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 KM SIGNAL.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1047 INTEGRIN ALPHA V SUBUNIT.
 SQ SEQUENCE 1047 AA; 116093 MW; 285F75F17471CA4B CRC64;

Query Match 43.9%; Score 50; DB 6; Length 1047;
 Best Local Similarity 38.1%; Pred. No. 98;
 Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFPAAKKIPEEKREKLE 21
 Db 1016 RMGFFKVRPQEQEREDQ 1036

RESULT 12
 064016 PRELIMINARY; PRT; 236 AA.
 ID 064016
 AC 064016
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Putative lipoprotein.
 GN YOKB.
 OS Bacteriophage SPB2.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=66797;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.;
 RT "The complete nucleotide sequence of the Bacillus subtilis SPB2ac2
 RT prophage.";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.


```

DR EMBL AF020713; AAC12975.1; -.
KW Lipoprotein.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4FAAC CRC64;

Query Match 43.4%; Score 49.5; DB 9; Length 236;
Best Local Similarity 84.6%; Pred. No. 28;
Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 8 KKI-PEEEKREEK 19
DB 205 KKVSPPEEKREEK 217

RESULT 13
ID 032005 PRELIMINARY; PRT; 236 AA.
AC 032005;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE YOKB protein.
GN YOKB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RX MEDLINE=98044033; PubMed=9384377;
RA Azevedo V., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Bortis R., Boursier L., Brans A., Braun M., Britnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut L.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorle B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogasawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Potwilk S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandendri M., Vannier F., Vasseroiti A.,
RA Viari A., Wambolt R., Wedler E., Wedler H., Weltenegeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yamumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z59115; CAB14083.1; -.
KW Complete proteome.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4FAAC CRC64;

Query Match 43.4%; Score 49.5; DB 16; Length 236;
Best Local Similarity 84.6%; Pred. No. 28;
Matches 11; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

```

```

QY 8 KKI-PEEEKREEK 19
DB 205 KKVSPPEEKREEK 217

RESULT 14
ID 0961R8 PRELIMINARY; PRT; 635 AA.
AC 0961R8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CDNA FLJ25142 f1s, clone CBR07252, highly similar to Homo sapiens
DE putative transcription factor-like nuclear regulator (TFNR gene)
DE (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RX MEDLINE=98044033; PubMed=9384377;
RA Azevedo V., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Bortis R., Boursier L., Brans A., Braun M., Britnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Gallazzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut L.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Jorle B., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogasawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Potwilk S., Prescott A.M.,
RA Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccoti E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandendri M., Vannier F., Vasseroiti A.,
RA Viari A., Wambolt R., Wedler E., Wedler H., Weltenegeger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yamumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z59115; CAB14083.1; -.
KW Complete proteome.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4FAAC CRC64;

Query Match 43.4%; Score 49.5; DB 4; Length 635;
Best Local Similarity 57.9%; Pred. No. 72;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 4 FFAH--KKIPEEEKREEK 19
DB 346 FFAHLLQVLAEEKREKOK 364

RESULT 15
ID 09HAW1 PRELIMINARY; PRT; 1372 AA.
AC 09HAW1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE RNA polymerase III transcription initiation factor B' short.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RX MEDLINE=20496900; PubMed=11040218;
RA Schramm L., Pendergrast P.S., Sun Y., Hernandez N.;
RL "Different human TFIIB activities direct RNA polymerase III
RL transcription from TATA-containing and TATA-less promoters."
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
DR EMBL; AF298152; AAG30221.1; -.
DR InterPro; IPR000977; DNA_ligase.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; Myb_DNA_binding; 1.
DR SMART; SM00395; SANT; 1.

```

DR PROSITE; PS00697; DNA LIGASE A1; UNKNOWN 1.
KW DNA-binding; Initiation factor; Nuclear protein.
SQ SEQUENCE 1372 AA; 154605 MW; C7416FD179610F22 CRC64;

Query Match 43.4%; Score 49.5; DB 4; Length 1372;
Best Local Similarity 57.9%; Pred. No. 1.5e+02;
Matches 11; Conservative 3; Mismatches 2; Indels 3; Gaps 1;

QY 4 FFAH--KRIPEEKREK 19
||| | : ||| : :
Db 366 FFAHLLQKVLAEERKOK 384

Search completed: July 16, 2003, 07:55:30
Job time : 4.25879 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:55:38 ; Search time 1.59353 Seconds
(without alignments)
1639.577 Million cell updates/sec

Title: US-09-647-544-7

Perfect score: 114
Sequence: 1 KLGFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	44.7	856	10 US-09-815-242-11310	Sequence 11310, A
2	49	43.0	856	10 US-10-083-357-1275	Sequence 1275, Ap
3	49	43.0	856	10 US-09-815-242-11489	Sequence 11489, A
4	46	40.4	479	10 US-09-971-309-64	Sequence 64, Appl
5	44	38.6	19	9 US-10-101-487-67	Sequence 67, Appl
6	44	38.6	20	9 US-09-520-907B-24	Sequence 24, Appl
7	44	38.6	93	9 US-10-134-841-2	Sequence 2, Appl
8	44	38.6	122	9 US-10-106-698-6119	Sequence 6119, Ap
9	44	38.6	176	9 US-10-101-487-70	Sequence 70, Appl
10	44	38.6	179	9 US-10-101-487-46	Sequence 46, Appl
11	44	38.6	186	9 US-10-101-487-44	Sequence 44, Appl
12	44	38.6	187	9 US-10-101-487-50	Sequence 50, Appl
13	44	38.6	198	9 US-10-101-487-81	Sequence 81, Appl
14	44	38.6	191	9 US-10-101-487-42	Sequence 42, Appl
15	44	38.6	200	9 US-09-736-457-789	Sequence 789, App
16	44	38.6	200	9 US-10-101-487-53	Sequence 53, Appl
17	44	38.6	200	9 US-09-902-941-789	Sequence 789, App
18	44	38.6	200	9 US-09-849-626-789	Sequence 789, App
19	44	38.6	200	9 US-10-017-754-789	Sequence 789, App

20	44	38.6	207	9 US-09-736-457-1667	Sequence 1667, Ap
21	44	38.6	207	9 US-09-902-941-1667	Sequence 1667, Ap
22	44	38.6	207	9 US-09-902-941-1913	Sequence 1913, Ap
23	44	38.6	207	9 US-09-849-626-1667	Sequence 1667, Ap
24	44	38.6	207	9 US-09-849-626-1913	Sequence 1913, Ap
25	44	38.6	207	9 US-10-017-754-1667	Sequence 1667, Ap
26	44	38.6	207	9 US-10-017-754-1913	Sequence 1913, Ap
27	44	38.6	221	9 US-10-094-458A-9	Sequence 9, Appl
28	44	38.6	430	9 US-10-286-264-104	Sequence 104, App
29	44	38.6	430	9 US-10-094-458A-3	Sequence 3, Appl
30	44	38.6	430	9 US-10-094-458A-6	Sequence 6, Appl
31	44	38.6	430	9 US-10-295-403-34	Sequence 34, Appl
32	44	38.6	430	9 US-09-934-455-154	Sequence 154, App
33	44	38.6	461	9 US-10-234-432-24	Sequence 24, Appl
34	44	38.6	676	9 US-10-128-714-3166	Sequence 3166, Ap
35	44	38.6	676	9 US-10-128-714-8166	Sequence 8166, Ap
36	44	38.6	893	9 US-10-234-432-86	Sequence 86, Appl
37	43.5	38.2	1456	9 US-10-205-823-113	Sequence 113, App
38	43.5	38.2	1494	9 US-10-205-823-111	Sequence 111, App
39	43.5	38.2	1863	9 US-09-734-672-2	Sequence 2, Appl
40	43.5	38.2	1863	9 US-09-734-672-4	Sequence 4, Appl
41	43.5	38.2	1863	9 US-09-734-672-6	Sequence 6, Appl
42	43.5	38.2	1863	9 US-09-982-828-2	Sequence 2, Appl
43	43.5	38.2	1863	9 US-09-982-828-4	Sequence 4, Appl
44	43.5	38.2	1863	9 US-09-982-828-6	Sequence 6, Appl
45	43.5	38.2	1863	9 US-10-022-819-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-11310
Sequence 11310, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11310
LENGTH: 856
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11310
Query Match 44.7%; Score 51; DB 10; Length 856;
Best Local Similarity 45.0%; Pred. No. 61;

Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
QY 3 GFAHKKPEEKREKLEQ 22
DB 496 GEIYSKIPENKKKEBELQR 515

RESULT 2

US-10-083-357-1275
; Sequence 1275, Application US/10083357
; Publication No. US20030054370A1
; GENERAL INFORMATION:
; APPLICANT: Qiandong Zeng et al.
; TITLE OF INVENTION: Systemic Discovery of New Genes
; FILE REFERENCE: 032796-090
; CURRENT APPLICATION NUMBER: US/10/083,357
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 1346
; SEQ ID NO 1275
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1275

Query Match 43.0%; Score 49; DB 9; Length 413;
Best Local Similarity 56.2%; Pred. No. 52;
Matches 9; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 7 HKKPEEKREKLEQ 22
DB 373 NKRIROKEREKLE 388

RESULT 3

US-09-815-242-11489
; Sequence 11489, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11489
; LENGTH: 856
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-815-242-11489

Query Match 43.0%; Score 49; DB 10; Length 856;
Best Local Similarity 53.3%; Pred. No. 1,1e+02;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
QY 7 HKKPEEKREKLE 21
DB 500 YSKIPEKEREKLEQ 514

RESULT 4

US-09-971-309-64
; Sequence 64, Application US/09971309
; Patent No. US2002010675A1
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunobu
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-0494P
; CURRENT APPLICATION NUMBER: US/09/971,309
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 09/446,504
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 64
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-971-309-64

Query Match 40.4%; Score 46; DB 10; Length 479;
Best Local Similarity 75.0%; Pred. No. 1,6e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 11 PEEKREKLEQ 22
DB 451 PEEKREKEREK 462

RESULT 5

US-10-101-487-67
; Sequence 67, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOPOWIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER, JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 67
; LENGTH: 19
; TYPE: PRT

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-67

Query Match 38.6%; Score 44; DB 9; Length 19;
Best Local Similarity 57.1%; Pred. No. 9.5;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEKREKLEQ 22
|:|||||:|:|:
Db 2 KMPPEEEEEEE 15

RESULT 6
US-09-320-907B-24
; Sequence 24, Application US/09320907B
; Publication No. US20020198360A1
; GENERAL INFORMATION:
; APPLICANT: GINSBERG, MARK H.
; APPLICANT: PRAFF, MARTIN
; TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC DOMAINS OF
; TITLE OF INVENTION: TRANSMEMBRANE RECEPTORS
; FILE REFERENCE: SRI-0006
; CURRENT APPLICATION NUMBER: US/09/320,907B
; CURRENT FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 09/187,236
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-320-907B-24

Query Match 38.6%; Score 44; DB 9; Length 20;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 KUGFFAHKKIPEEKKEE 18
|:|||||:|:|:
Db 1 KVGFFKRNRPPLLEAAEE 18

RESULT 7
US-10-134-841-2
; Sequence 2, Application US/10134841
; Publication No. US20030003482A1
; GENERAL INFORMATION:
; APPLICANT: HALL, JOHN-PETER
; APPLICANT: GOSPELT, ANDREAS
; TITLE OF INVENTION: MRP8/MRP14 heterodimer, or its
; TITLE OF INVENTION: individual components in combination, for treating and/or
; TITLE OF INVENTION: preventing skin diseases, wounds and/or wound-healing
; TITLE OF INVENTION: disturbances, having a reduced quantity of MRP8/MRP14
; TITLE OF INVENTION: heterodimers
; FILE REFERENCE: 50125/031002
; CURRENT APPLICATION NUMBER: US/10/134,841
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: US 60/332,925
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: DE 10121254.2
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-134-841-2

Query Match 38.6%; Score 44; DB 9; Length 93;
Best Local Similarity 52.9%; Pred. No. 51;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 KUGFFAHKKIPEEKKEE 17
|:|||||:|:|:
Db 77 KMGVAHKKSHESHKE 93

RESULT 8
US-10-106-698-6119
; Sequence 6119, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6119
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6119

Query Match 38.6%; Score 44; DB 9; Length 122;
Best Local Similarity 47.1%; Pred. No. 68;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPEEKREKLEQ 22
|:|||||:|:|:
Db 98 ARKKVEEDDEDEDEEE 114

RESULT 9
US-10-101-487-70
; Sequence 70, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOULIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-70

Query Match 38.6%; Score 44; DB 9; Length 176;
Best Local Similarity 57.1%; Pred. No. 1e+02;

Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Qy 9 KIPPEEKREKLEQ 22
Db 2 KMPEEEEEEEEEEE 15

RESULT 10
US-10-101-487-46
; Sequence 46, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUJIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-46

Query Match 38.6%; Score 44; DB 9; Length 179;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPPEEKREKLEQ 22
Db 2 KMPEEEEEEEEEEE 15

RESULT 11
US-10-101-487-44
; Sequence 44, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUJIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion
US-10-101-487-44

Query Match 38.6%; Score 44; DB 9; Length 186;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPPEEKREKLEQ 22
Db 2 KMPEEEEEEEEEEE 22

RESULT 12
US-10-101-487-50
; Sequence 50, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUJIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-50

Query Match 38.6%; Score 44; DB 9; Length 187;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 9 KIPPEEKREKLEQ 22
Db 2 KMPEEEEEEEEEEE 15

RESULT 13
US-10-101-487-81
; Sequence 81, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUJIST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 81
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-81

Query Match 38.6%; Score 44; DB 9; Length 191;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEEKREKLEQ 22
|:||||:|:|:
Db 2 KMPEEEEEEEEE 15

RESULT 14

US-10-101-487-42
; Sequence 42; Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
; APPLICANT: LEUNG, DAVID W.
; APPLICANT: BERGMAN, PHILIP A.
; APPLICANT: LOFOUISST, ALAN
; APPLICANT: PIETZ, GREGORY E.
; APPLICANT: TOMPKINS, CHRISTOPHER K.
; APPLICANT: WAGGONER, JR., DAVID W.
; TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
; FILE REFERENCE: 077319/0329
; CURRENT APPLICATION NUMBER: US/10/101,487
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/277,705
; PRIOR FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic fusion
; OTHER INFORMATION: protein
US-10-101-487-42

Query Match 38.6%; Score 44; DB 9; Length 198;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEEKREKLEQ 22
|:||||:|:|:
Db 9 KMPEEEEEEEEE 22

RESULT 15

US-09-736-457-789
; Sequence 789; Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 789
; LENGTH: 200
; TYPE: PRT

; ORGANISM: Homo sapien
US-09-736-457-789

Query Match 38.6%; Score 44; DB 9; Length 200;
Best Local Similarity 47.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPPEEKREKLEQ 22
|:||||:|:|:
Db 176 ARKKVEEEDDEEEEEEE 192

Search completed: July 16, 2003, 08:12:53
Job time : 2.59353 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:45:22 ; Search time 2.44444 Seconds
(without alignments)
1199.256 Million cell updates/sec

Title: US-09-647-544-7
Page: 114

Sequence: 1 KLGFFAHKKIP EE EKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

1: A_Geneseq.101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query % Match	length	DB	ID	Description
1	114	100.0	22	21	AAV32244	Human integrin sub	
2	114	100.0	1132	21	AAV32243	Human integrin sub	
3	114	100.0	1167	21	AAV32242	Human integrin sub	
4	114	100.0	1167	22	ABG64584	Human secreted p	
5	51	44.7	121	22	ABG26508	Novel human diag	
6	51	44.7	147	22	ABG11269	Novel human diag	
7	51	44.7	502	22	ABG07742	Novel human diag	
8	51	44.7	856	22	AAU35717	Novel human diag	
9	50	43.9	32	16	AAR80855	Helicobacter pylor	
10	50	43.9	156	22	ABG18894	Integrin alpha-V c	
						Novel human diag	

[illegible]

ALIGNMENTS

XX	AAV32244	RESULT 1
ID	AAV32244 strand; Peptide: 22 AA.	
XX	AAV32244;	
AC		
DT	15-FEB-2000 (first entry)	
XX		
DE	Human integrin subunit alpha-10 cytoplasmic domain peptide.	
XX		
KW	Integrin alpha-10; Isa10; human; trauma; rheumatoid arthritis;	
KW	osteoarthritis; osteoarthritis; cancer; atherosclerosis;	
KW	inflammation; therapy; cartilage; chondrocyte; osteoblast;	
KW	fibroblast; vaccine; marker.	
XX		
OS	Homo sapiens.	
XX		
FN	W09951639-A1;	
XX		
PD	14-OCT-1999.	
XX		
PF	31-MAR-1999; 99WO-SE00544.	
XX		
PR	02-APR-1998; 98SE-0001164.	
XX	28-JAN-1999; 99SE-0000319.	
PA	(ACTI-) ACTIVE BIOTECH AB.	
XX		
PI	Lundgren-Akerlund E;	
XX		
DR	WPI; 2000-052639/04.	
PT	New isolated integrin subunit alpha-10, used as a marker or target	

PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation -

PS Claim 21; Page 53; 90pp; English.

CC This sequence represents a fragment of novel human chondrocyte
 CC integrin subunit alpha-10 (Isa10, see AAY32242), corresponding to
 CC the C-terminal cytoplasmic domain of the protein. The invention
 CC relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The integrin heterodimer, or the subunit
 CC alpha-10, or a fragment of it such as the present sequence, can be
 CC used as a marker or target of all types of cells, e.g. of
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:
 CC for treating pathological conditions involving Isa10, such as
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage,
 CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). Isa10 binding entities can be used to
 CC determine the differentiation-state of cells during embryonic
 CC development, angiogenesis or development of cancer, in pathological
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
 CC in tissue regeneration or in therapeutic and physiological repair
 CC of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. Isa10
 CC polynucleotides, vectors, host cells and methods of producing
 CC recombinant Isa10 are also claimed.

SQ Sequence 22 AA;

Query Match 100.0%; Score 114; DB 21; Length 22;
 Best Local Similarity 100.0%; Pred. No. 2.7e-09;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFFAHKKIPEEKREKLEQ 22
 |||||
 DB 1 KLGFFAHKKIPEEKREKLEQ 22

RESULT 2
 AAY32243
 ID AAY32243 standard; Protein; 1132 AA.

AC AAY32243;
 XX
 DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 splice variant.

DE Integrin alpha-10; Isa10; human; trauma; rheumatoid arthritis;
 XX
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KW fibroblast; vaccine; marker; splice variant.

XX Homo sapiens.

OS
 XX
 XX Key Location/Qualifiers
 FH 1..22
 FT Peptide /note= "signal peptide"
 FT Protein 23..1132
 FT /note= "mature protein"

XX WO951639-A1.
 XX
 PD 14-OCT-1999.

PF 31-MAR-1999; 99WO-SE00544.

XX 02-APR-1998; 98SE-0001164.

PR 28-JAN-1999; 99SE-0000319.

XX (ACTI-) ACTIVE BIOTECH AB.

PA Lundgren-Akerlund E;

PI WPI: 2000-052639/04.

XX N-PSDB; AA234720.

DR New isolated integrin subunit alpha-10, used as a marker or target

PT molecule for cells during development, regeneration and pathological

PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or

PT inflammation -

PS Claim 1; Page 43-48; 90pp; English.

CC This sequence represents a splice variant of novel human
 CC chondrocyte integrin subunit alpha-10 (Isa10). It is identical to
 CC Isa10 (see AAY32242) except for deletion of amino acids 975-986. The
 CC invention relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The heterodimer, subunit alpha-10 or splice
 CC variant can be used as a marker or target of all types of cells, e.g.
 CC of chondrocytes, osteoblasts and fibroblasts. They can also be used:
 CC for treating pathological conditions involving Isa10, such as
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic repair of cartilage,
 CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). Isa10 binding entities can be used to
 CC determine the differentiation-state of cells during embryonic
 CC development, angiogenesis or development of cancer, in pathological
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
 CC in tissue regeneration or in therapeutic and physiological repair
 CC of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. Isa10
 CC polynucleotides, vectors, host cells and methods of producing
 CC recombinant Isa10 are also claimed.

SQ Sequence 1132 AA;

Query Match 100.0%; Score 114; DB 21; Length 1132;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFFAHKKIPEEKREKLEQ 22
 |||||
 DB 1111 KLGFFAHKKIPEEKREKLEQ 1132

RESULT 3
 AAY32242
 ID AAY32242 standard; Protein; 1167 AA.

AC AAY32242;

XX 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10.

DE Integrin alpha-10; Isa10; human; trauma; rheumatoid arthritis;
 XX
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KW fibroblast; vaccine; marker.

OS Homo sapiens.
 XX Key
 FH Location/Qualifiers
 FT 1..22
 FT /note= "signal peptide"
 FT 23..1145
 FT /note= "mature protein"
 FT 23..1120
 FT /note= "extracellular domain"
 FT 1121..1145
 FT /note= "transmembrane domain"
 FT 1122..1167
 FT /note= "cytoplasmic domain, specifically claimed
 in Claim 21"
 FT 162..359
 FT /note= "I-domain"
 FT 494..502
 FT /note= "cation binding site motif"
 FT 558..566
 FT /note= "cation binding site motif"
 FT 620..628
 FT /note= "cation binding site motif"
 FT 98
 FT /note= "N-glycosylated"
 FT 336
 FT /note= "N-glycosylated"
 FT 364
 FT /note= "N-glycosylated"
 FT 733
 FT /note= "N-glycosylated"
 FT 839
 FT /note= "N-glycosylated"
 FT 921
 FT /note= "N-glycosylated"
 FT 1018
 FT /note= "N-glycosylated"
 FT 1039
 FT /note= "N-glycosylated"
 FT /note= "N-glycosylated"
 FT MO9951639-A1.
 PN 14-OCT-1999.
 PD 31-MAR-1999; 99WO-SE00544.
 XX 02-APR-1998; 98SE-0001164.
 PR 28-JAN-1999; 99SE-0000319.
 XX (ACT1-) ACTIVE BIOTECH AB.
 PA Lundgren-Akerlund E;
 PI WPI, 2000-052639/04.
 DR N-PSDB; AA234719.
 XX New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation -
 PS Claim 1; Fig 6; 90pp; English.
 XX This sequence represents novel human chondrocyte integrin subunit
 CC alpha-10 (ISa10). A splice variant is given in AA32243. The
 CC invention relates to a recombinant or isolated integrin heterodimer
 CC comprising the alpha10 subunit in association with subunit beta
 CC (especially beta-1). The heterodimer and the subunit alpha-10 can
 CC be used as markers or targets of all types of cells, e.g. of
 CC chondrocytes, osteoblasts and fibroblasts. They can also be used:
 CC for treating pathological conditions involving ISa10, such as
 CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
 CC for detecting the formation of cartilage during embryonal
 CC development, physiological or therapeutic reparation of cartilage,

CC or detecting regeneration of cartilage or chondrocytes during
 CC transplantation of cartilage or chondrocytes; for selection and
 CC analysis or for sorting, isolating or purification of chondrocytes
 CC and for in vitro studies of differentiation of chondrocytes; and as
 CC a target for anti-adhesive drugs or molecules in tendon, ligament,
 CC skeletal muscle or other tissues where adhesion impairs the function
 CC of the tissue (all claimed). ISa10 binding entities can be used to
 CC determine the differentiation-state of cells during embryonic
 CC development, angiogenesis or development of cancer, in pathological
 CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
 CC in tissue regeneration or in therapeutic and physiological reparation
 CC of cartilage (claimed). A vaccine comprising the integrin
 CC heterodimer or subunit alpha-10 is also claimed. ISa10
 CC polynucleotides, vectors, host cells and methods of producing
 CC recombinant ISa10 are also claimed.
 XX Sequence 1167 AA;
 SQ
 Query Match 100.0%; Score 114; DB 21; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGFPAHKKIPBEKREKLEQ 22
 DB 1146 KLGFPAHKKIPBEKREKLEQ 1167
 RESULT 4
 ID AAB64584 standard; Protein; 1167 AA.
 XX AAB64584;
 AC 22-MAR-2001 (first entry)
 XX
 DT Human secreted protein #37.
 XX
 DE Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 XX vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
 XX cardiatic; gene therapy; cancer; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; human; secreted protein.
 OS Homo sapiens.
 XX WO200077197-A1.
 PN 21-DEC-2000.
 PD 01-JUN-2000; 2000WO-US14934.
 XX 11-JUN-1999; 99US-0138599.
 PR (HDMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX Rosen CA, Ruben SM, Komatsoulis GA;
 PI WPI, 2001-032312/04.
 DR N-PSDB; AAF32793.
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 PS Claim 11; Page 496-500; 558pp; English.
 XX Sequences AAB64549-B64594 represent the amino acid sequences of 47
 CC human secreted proteins encoded by the genes AAF32757-F32803. The genes
 CC and proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from
 CC a range of human tissues disclosed in the specification. The nucleic
 CC acids, proteins, antibodies and (ant)agonists are useful in the
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.

XX
 SQ Sequence 1167 AA;
 Query Match 100.0%; Score 114; DB 22; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 1.4e-07;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGFPAHKKIPPEEKREKLEQ 22
 |||||
 DB 1146 KLGFPAHKKIPPEEKREKLEQ 1167

RESULT 5
 ABG26508
 ID ABG26508 standard; Protein; 121 AA.
 XX
 AC ABG26508;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #26499.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Dmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS90695.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID No 56867; 103pp; English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC responsible for genetic disorders or other traits to assess biodiversity,
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPD
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 121 AA;
 Query Match 44.7%; Score 51; DB 22; Length 121;
 Best Local Similarity 55.6%; Pred. No. 10;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKLEQ 22
 |||||
 DB 5 FLVKKIKKEEEEEEEEEE 22

RESULT 6
 ABG11269
 ID ABG11269 standard; Protein; 147 AA.
 XX
 AC ABG11269;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #11260.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Dmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS75456.
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 20; SEQ ID No 41628; 103pp; English.

XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

ABBA48958
ID ABB48958 standard; Protein; 357 AA.
AC ABB48958;
XX
XX
XX 05-FEB-2002 (first entry)
XX
XX
XX Listeria monocytogenes protein #1662.
DE
XX
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
XX Listeria monocytogenes.
OS
XX
XX WO200177335-A2.
XX
XX 18-OCT-2001.
XX
XX 11-APR-2001; 2001WO-FR01118.
XX
XX 11-APR-2000; 2000FR-0004629.
XX
XX (INSP) INST PASTEUR.
XX
XX Buchrieser C, Frangeul L, Couve E, Rusnick C, Fsihi H, Dehoux P,
PI Dussurget O, Chetoui F, Nedjari H, Glaeser P, Kunst F, Coissart P,
PI Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA,
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
PI Chakraborty T, Domann E, Hain T, Berche P, Chablit A, Durant L,
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
XX WPI; 2002-010914/01.
XX
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
XX
XX Claim 6; SEQ ID No 1663; 192pp; French.
XX
XX
XX The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABB403041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies. Identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 357 AA;
SQ
Query Match 43.9%; Score 50; DB 23; Length 357;
Best Local Similarity 71.4%; Pred. No. 42;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 6 AHKXIPPEEKREK 19
DB 146 AKKRTPEEKREK 159

AAU76335 standard; peptide; 1048 AA.
ID AAU76335;
XX
XX
XX
XX 21-MAY-2002 (first entry)
XX
XX
XX Human anti-dual integrin protein #1.
XX
XX
XX Human; dual integrin; HC CDR; variable region;
KW LC CDR; medical device; immune related disease; rheumatoid arthritis;
KW gastric ulcer; asthma; allergic rhinitis; Crohn's pathology; atherosclerosis;
KW sickle cell anaemia; diabetes; cardiovascular disease; arteriosclerosis;
KW atherosclerosis; restenosis; angina pectoris; myocardial infarction;
KW infectious disease; pneumonia; leprosy; malaria; malignant disease;
KW leukaemia; chronic myelocytic leukaemia; multiple myeloma;
KW neurological disease; multiple sclerosis; Parkinson's disease;
KW Alzheimer's disease; Creutzfeldt-Jakob disease.
XX
XX
XX Homo sapiens.
OS
XX
XX WO200212501-A2.
XX
XX 14-FEB-2002.
XX
XX 07-AUG-2001; 2001WO-US24784.
XX
XX 07-AUG-2000; 2000US-223363P.
XX
XX 01-AUG-2001; 2001US-0920267.
XX
XX (CENZ) CEN2OCOR INC.
XX
XX Giles-Komar J, Heavner G, Snyder L, Trikha M;
PI
XX
XX WPI; 2002-217193/27.
XX
XX
XX Novel isolated mammalian anti-dual integrin antibody, useful for
PT diagnosing or treating dual integrin related condition such as
PT rheumatoid arthritis, gastric ulcer, asthma, atherosclerosis,
PT restenosis -
XX
XX
XX Claim 81; Page 135-138; 144pp; English.
XX
XX
XX The invention relates to an isolated mammalian anti-dual integrin
CC antibody having at least one of the human heavy chain or light
CC chain complementary determining region (CDR, HC CDR1-CDR3, LC
CC CDR1-3). Also included are the nucleic acids encoding the CDRs,
CC a vector comprising the nucleic acids, a host cell comprising the
CC vector, an anti-idiotypic antibody that binds to the anti-dual
CC integrin, a medical device comprising the antibody suitable for
CC administration by parenteral, subcutaneous, intramuscular, intravenous,
CC intracardiac, intrabronchial, intrabdominal, intracapsular,
CC intracartilaginous, intracavitary, intracerebellar, or other
CC routes as given in specification. The antibody is useful for diagnosing
CC or treating a dual integrin related condition in an animal for example,
CC immune related disease such as rheumatoid arthritis, gastric ulcer,
CC asthma, allergic rhinitis, Crohn's pathology, sickle cell anaemia,
CC diabetes, cardiovascular disease such as arteriosclerosis,
CC atherosclerosis, restenosis, angina pectoris, myocardial infarction,
CC infectious disease in a cell such as bacterial, viral, and fungal
CC infections, pneumonia, leprosy, malaria, malignant disease such as
CC leukaemia, chronic myelocytic leukaemia, Burkitt's lymphoma, multiple
CC myeloma; neurological disease such as multiple sclerosis, Parkinson's
CC disease, spinal ataxia, Alzheimer's disease, Creutzfeldt-Jakob
CC disease and many other diseases given in the specification. The
CC present sequence is an human dual integrin protein against which the
CC antibodies of the invention were raised.
XX
XX
XX Sequence 1048 AA;
SQ
Query Match 43.9%; Score 50; DB 23; Length 1048;
Best Local Similarity 38.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 43.0%; Score 49; DB 21; Length 486;
Best Local Similarity 71.4%; Pred.No. 78;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 8 KKIPEEKREKLE 21
||| | :||| :|||
Db 267 KKIPEEKREKKE 280

Search completed: July 16, 2003, 07:52:08
Job time : 3.44444 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:49:17 ; Search time 0.866385 Seconds
(without alignments)
747.132 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFPAHKKIPDEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	43.9	32	1	US-08-214-770-2
2	50	43.9	32	5	PCT-US95-02885-2
3	48	42.1	573	2	US-08-745-934-4
4	48	42.1	573	4	US-09-147-009-7
5	47.5	41.7	2343	4	US-09-324-867-2
6	46	40.4	479	4	US-09-446-504-64
7	46	40.4	479	4	US-09-712-266-64
8	45	39.5	109	4	US-09-091-725-31
9	44	38.6	46	1	US-08-200-016-3
10	44	38.6	93	1	US-07-987-272A-7
11	44	38.6	93	1	US-07-987-272A-16
12	44	38.6	93	1	US-08-385-241-1
13	43.5	38.2	765	1	US-08-425-061-19
14	43.5	38.2	765	2	US-08-825-886-19
15	43.5	38.2	900	1	US-08-425-061-20
16	43.5	38.2	900	2	US-08-825-886-20
17	43.5	38.2	914	1	US-08-425-061-21
18	43.5	38.2	914	2	US-08-825-886-21
19	43.5	38.2	1202	1	US-08-425-061-22
20	43.5	38.2	1202	2	US-08-825-886-22
21	43.5	38.2	1363	1	US-08-425-061-23
22	43.5	38.2	1363	2	US-08-825-886-23
23	43.5	38.2	1852	1	US-08-425-061-24
24	43.5	38.2	1852	2	US-08-825-886-24
25	43.5	38.2	1863	1	US-08-425-061-16
26	43.5	38.2	1863	1	US-08-598-591-2
27	43.5	38.2	1863	1	US-08-480-784-2

28	43.5	38.2	1863	1	US-08-483-553-2	Sequence 2, Appli
29	43.5	38.2	1863	1	US-08-487-002-2	Sequence 2, Appli
30	43.5	38.2	1863	1	US-08-483-554B-2	Sequence 2, Appli
31	43.5	38.2	1863	1	US-08-798-691-2	Sequence 2, Appli
32	43.5	38.2	1863	1	US-08-798-691-4	Sequence 2, Appli
33	43.5	38.2	1863	1	US-08-798-691-6	Sequence 6, Appli
34	43.5	38.2	1863	1	US-08-488-011B-2	Sequence 2, Appli
35	43.5	38.2	1863	2	US-08-825-886-16	Sequence 16, Appli
36	43.5	38.2	1863	2	US-08-603-753D-2	Sequence 2, Appli
37	43.5	38.2	1863	3	US-08-825-487A-2	Sequence 2, Appli
38	43.5	38.2	1863	3	US-08-825-487A-4	Sequence 4, Appli
39	43.5	38.2	1863	3	US-08-825-487A-6	Sequence 6, Appli
40	43.5	38.2	1863	4	US-09-074-476-2	Sequence 2, Appli
41	43.5	38.2	1863	4	US-09-074-476-4	Sequence 4, Appli
42	43.5	38.2	1863	4	US-09-074-476-6	Sequence 6, Appli
43	43.5	38.2	1863	4	US-09-099-753-2	Sequence 2, Appli
44	43.5	38.2	1863	4	US-08-850-172-2	Sequence 2, Appli
45	43.5	38.2	1863	4	US-08-986-106-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-214-770-2
; Sequence 2, Application US/08214770
; Patent No. 5523209
; GENERAL INFORMATION:
; APPLICANT: Ginsberg, Mark H.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: INHIBITORS OF INTEGRIN
; TITLE OF INVENTION: ACTIVATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,770
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06410/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-214-770-2

Query Match 43.9%; Score 50; DB 1; Length 32;
Best Local Similarity 38.1%; Pred. No. 1.2;
Matches 8; Conservative 7; Mismatches 0; Gaps 0;

Qy 1 KLGFFAHKKIPPEEKREKLE 21
:||||:||||:||||:
Db 1 RMGFFKRVPRPOEORREQLQ 21

RESULT 2

PCT-US95-02885-2
Sequence 2, Application PC/RUS9502885
GENERAL INFORMATION:
APPLICANT: Ginsberg, Mark H.
APPLICANT: O'Toole, Timothy
TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS
TITLE OF INVENTION: OF INTEGRIN ACTIVATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02885
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/214,770
FILING DATE: March 14, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06410/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 32
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US95-02885-2

Query Match 43.9%; Score 50; DB 5; Length 32;
Best Local Similarity 38.1%; Pred. No. 1.2;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KLGFFAHKKIPPEEKREKLE 21
:||||:||||:||||:
Db 1 RMGFFKRVPRPOEORREQLQ 21

RESULT 3
US-08-745-934-4
Sequence 4, Application US/08745934
Patent No. 5861496
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Hawkins, Philip R.
TITLE OF INVENTION: HUMAN SQUALENE EPOXIDASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: California
COUNTRY: USA

ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/745,934
FILING DATE: Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0151 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 855-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 1083804
US-08-745-934-4

Query Match 42.1%; Score 48; DB 2; Length 573;
Best Local Similarity 52.9%; Pred. No. 38;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LGFPAHKKIPPEEKREK 18
:||||:||||:||||:
Db 75 IGFPAKSPPESEKKEQ 91

RESULT 4
US-09-147-009-7
Sequence 7, Application US/09147009
Patent No. 6153815
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Process for Raising Squalene Levels in Plants
TITLE OF INVENTION: and DNA Sequences Used Therefor
NUMBER OF SEQUENCES: 11
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/147,009
FILING DATE:
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 573 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Rattus norvegicus
TISSUE TYPE: kidney
CELL LINE: NRK
IMMEDIATE SOURCE:
LIBRARY: pcdb library of H. Okayama
CLONE: Tb-1

```

CURRENT FILING DATE: 1999-12-23
PRIORITY APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 479
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-446-504-64

Query Match          40.4%; Score 46; DB 4; Length 479;
Best Local Similarity 75.0%; Pred. No. 60;
Matches          9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      11 PEEKEEKKLEQ 22
      ||||| ||| :
      451 PEEKEEKKKEX 462

Db

RESULT 7
US-09-712-266-64
Sequence 64, Application US/09712266
Patent No. 633158
GENERAL INFORMATION:
APPLICANT: UEMORI, Takashi
APPLICANT: SATO, Yoshimi
APPLICANT: FUJITA, Tomoko
APPLICANT: MIYAKE, Kazuo
APPLICANT: MUKAI, Hiroyuki
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
FILE REFERENCE: 1422-408PCT
CURRENT APPLICATION NUMBER: US/09/712,266
CURRENT FILING DATE: 2000-11-15
PRIOR APPLICATION NUMBER: US 09/446,504
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 479
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-712-266-64

Query Match          40.4%; Score 46; DB 4; Length 479;
Best Local Similarity 75.0%; Pred. No. 60;
Matches          9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      11 PEEKEEKKLEQ 22
      ||||| ||| :
      451 PEEKEEKKKEX 462

Db

RESULT 8
US-09-091-725-31
Sequence 31, Application US/09091725
Patent No. 6329141
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Improved methods for transforming Phaffia
and recombinant DNA for use therein

```

NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster LLP
STREET: 2000 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,725
FILING DATE: 23-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95203620.0
FILING DATE: 22-DEC-1995
APPLICATION NUMBER: EP 96200943.7
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: E. Victor Donahue
REGISTRATION NUMBER: 35,492
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 109 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-091-725-31

Query Match 39.5%; Score 45; DB 4; Length 109;
Best Local Similarity 52.6%; Pred. No. 19;
Matches 10; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

3 GFFAHKIPPEEKRE 21
76 GASADASAPAEKKEKAE 94

RESULT 9
US-08-200-016-3
Sequence 3, Application US/08200016
Patent No. 5614397
GENERAL INFORMATION:
APPLICANT: Weisman, Irving
ADDRESSEE: Lagasse, Eric
TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR MODULATING
TITLE OF INVENTION: APOPTOSIS IN HEMATOLOGY CELLS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/200,016
FILING DATE: 22-FEB-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: 06037/003001
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 854-5277
TELEFAX: (415) 854-0875
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 46 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-200-016-3

Query Match 38.6%; Score 44; DB 1; Length 46;
Best Local Similarity 52.9%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

1 KLGFAHKKIPPEEKRE 17
30 KMGVAHKKSHESKHE 46

RESULT 10
US-07-987-272A-7
Sequence 7, Application US/07987272A
Patent No. 5731166
GENERAL INFORMATION:
APPLICANT: Geczy, C., Simpson, R. J. and Lackmann, M
TITLE OF INVENTION: No. 5731166el Chemotactic Factor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman Darby & Cushman
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/987,272A
FILING DATE: 05-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 2127
FILING DATE: 05-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 4463
FILING DATE: 05-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Brinkman, David W
REGISTRATION NUMBER: 20,817
REFERENCE/DOCKET NUMBER: DWB/1925/200259
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861 3000
TELEFAX: 202-822 0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-987-272A-7

Query Match 38.6%; Score 44; DB 1; Length 93;
Best Local Similarity 52.9%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

1 KLGFAHKKIPPEEKRE 17
77 KMGVAHKKSHESKHE 93

RESULT 11
US-07-987-272A-16
Sequence 16, Application US/07987272A
Patent No. 573166
GENERAL INFORMATION:
APPLICANT: Geczy, C., Simpson, R. J. and Lackmann, M
TITLE OF INVENTION: No. 573166el Chemotactic Factor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cushman Darby & Cushman
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/987,272A
FILING DATE: 05-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 2127
FILING DATE: 05-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 4463
FILING DATE: 05-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Brinkman, David W
REGISTRATION/DOCKET NUMBER: 20,817
REFERENCE/DOCKET NUMBER: DMB/1925/200259
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861 3000
TELEFAX: 202-822 0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-987-272A-16
Query Match 38.6%; Score 44; DB 1; Length 93;
Best Local Similarity 52.9%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 KLGFPAAKKIPEEKRE 17
Db 77 KMGVAHHKSHESHSKE 93
RESULT 12
US-08-385-241-1
Sequence 1, Application US/08385241
Patent No. 5776348
GENERAL INFORMATION:
APPLICANT: Selengut Ph.D., Jeremy D.
APPLICANT: Orme-Johnson Ph.D., William H.
APPLICANT: Dreidler M.D., Stephen P.
APPLICANT: Asakura M.D., Hirokaka
TITLE OF INVENTION: SYSTEM AND METHOD FOR INHIBITING
FORMATION OF CRYSTALLINE STRUCTURES THAT INCLUDE STRUVITE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Choate Hall & Stewart
STREET: 53 State Street
CITY: Boston

STATE: MA
COUNTRY: USA
ZIP: 02109-2891
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/385,241
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Herschbach Ph.D., Brenda M.
REGISTRATION NUMBER: P-39,223
REFERENCE/DOCKET NUMBER: 492611-000 (MIT6915)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-5175
TELEFAX: (617) 248-4000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: hmrp-8 protein
US-08-385-241-1

Query Match 38.6%; Score 44; DB 1; Length 93;
Best Local Similarity 52.9%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 KLGFPAAKKIPEEKRE 17
Db 77 KMGVAHHKSHESHSKE 93

RESULT 13
US-08-425-061-19
Sequence 19, Application US/08425061
Patent No. 5622829
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
CANCER
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,061
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-061-19

Query Match 38.2%; Score 43.5; DB 1; Length 765;
Best Local Similarity 58.8%; Pred. No. 2.1e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKPIPEEKREKLE 21
| : : ||| ||| |||
DB 721 FVNPSLPREK-EKLE 736

RESULT 14
US-08-825-886-19
Sequence 19, Application US/08825886
Patent No. 5821328
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825, 886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/425, 061
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 765 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-825-886-19

Query Match 38.2%; Score 43.5; DB 2; Length 765;
Best Local Similarity 58.8%; Pred. No. 2.1e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKPIPEEKREKLE 21
| : : ||| ||| |||
DB 721 FVNPSLPREK-EKLE 736

RESULT 15
US-08-425-061-20
Sequence 20, Application US/08425061
Patent No. 5622829
GENERAL INFORMATION:
APPLICANT: KING, Mary-Claire
APPLICANT: FRIEDMAN, Lori
APPLICANT: OSTERMEYER, Beth
APPLICANT: ROWELL, Sarah
APPLICANT: LYNCH, Eric
APPLICANT: SZABO, Csilla
APPLICANT: LEE, Ming
TITLE OF INVENTION: GENETIC MARKERS FOR BREAST AND OVARIAN
TITLE OF INVENTION: CANCER
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425, 061
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59563-3/DJB/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 900 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-061-20

Query Match 38.2%; Score 43.5; DB 1; Length 900;
Best Local Similarity 58.8%; Pred. No. 2.4e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 5 FAHKPIPEEKREKLE 21
| : : ||| ||| |||
DB 721 FVNPSLPREK-EKLE 736

Search completed: July 16, 2003, 07:57:49
Job time: 1.86639 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:11:08 ; Search time 15 Seconds
(without alignments)
140.997 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFPAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 4263

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	35.1	19	2	A49725
2	30	26.3	20	2	I53671
3	29	25.4	17	2	S10786
4	29	25.4	20	2	S43627
5	28	24.6	10	2	S65385
6	27	23.7	8	2	I64832
7	27	23.7	15	2	PA0041
8	27	23.7	16	2	S13898
9	27	23.7	17	2	G83975
10	27	23.7	18	2	A61220
11	27	23.7	20	2	S11416
12	26	22.8	15	2	S36891
13	26	22.8	18	2	E33178
14	26	22.8	21	2	PC7043
15	26	22.8	22	2	P33084
16	26	22.8	22	2	PC4388
17	25	21.9	16	2	A35552
18	25	21.9	16	2	JH0517
19	25	21.9	18	2	S39845
20	25	21.9	20	2	D49164
21	25	21.9	20	2	S06149
22	25	21.9	20	2	S77983
23	25	21.9	21	2	PK0078
24	24.5	21.5	22	2	SX00189
25	24	21.1	14	2	S65392
26	24	21.1	18	2	B44995
27	24	21.1	19	2	A37968
28	24	21.1	22	2	H86433
29	23.5	20.6	19	2	S02808

ALIGNMENTS

30	23	20.2	14	2	A49018	myosin heavy chain
31	23	20.2	15	2	C37765	hypothetical prote
32	23	20.2	18	2	I40062	shikimate 5-dehydr
33	23	20.2	19	2	S69166	ferridoxin b - Jap
34	23	20.2	20	2	JP0070	ribosomal protein
35	23	20.2	20	2	S06150	photosystem I chai
36	23	20.2	20	2	S19618	globin - polychaet
37	23	20.2	20	2	A31516	lectin, galactose/
38	23	20.2	21	2	S78416	ribosomal protein
39	23	20.2	22	2	JP0071	ribosomal protein
40	23	20.2	22	2	CE4330	ribosomal protein
41	23	20.2	22	2	S48136	lucosyltransferase
42	23	20.2	22	2	S78007	lucosyltransferase
43	22.5	19.7	20	2	A05310	apolipoprotein B -
44	22	19.3	10	2	S43625	cytochrome-c oxida
45	22	19.3	12	2	S16335	beta-conglycinin a

RESULT 1

A49725
valine-tRNA ligase (EC 6.1.1.9) - rabbit (fragment)
N.Alternate names: valyl-tRNA synthetase
C.Species: Oryctolagus cuniculus (domestic rabbit)
C.Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 03-Jun-2002
C.Accession: A49725
R.Bec, G., Kerian, P., Waller, J.P.
J. Biol. Chem. 269, 2086-2092, 1994
A.Title: Reconstitution in Vitro of the valyl-tRNA synthetase-elongation factor (EF) 1b
A.Reference number: A49725; MUID:94124563; PMID:8294461
A.Accession: A49725
A.Status: preliminary
A.Molecule type: protein
A.Residues: 1-19 <BEC>
C.Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 35.1%; Score 40; DB 2; Length 19;

Best local similarity 50.0%; Pred. No. 26;

Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLEQ 22

Db 2 QLPXKAKNEKLEK 15

RESULT 2

I53671
neurofilament heavy subunit - human (fragment)

C.Species: Homo sapiens (man)
C.Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C.Accession: I53671
R.Figlewicz, D.A.; Rouleau, G.A.; Krizus, A.; Jullien, J.P.

A.Title: Polymorphism in the multi-phosphorylation domain of the human neurofilament hea

A.Reference number: I53671; MUID:94040777; PMID:8224877

A.Accession: I53671
A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: mRNA
A.Residues: 1-20 <RES>

A.Cross-references: GB:S66488; NID:G452861; PIDN:AA828609.1; PID:G452862

C.Genetics:
A.Gene: GDB:NEFH

A.Cross-references: GDB:120225; OMIM:162230

A.Map position: 22q12.1-22q13.1
Query Match 26.3%; Score 30; DB 2; Length 20;

Best local similarity 53.8%; Pred. No. 6.5e+02;

Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 9 KIPEEKREKLE 21

Db 6 KSPKAKSPKPE 18

RESULT 3

enamelin, 26k - bovine (fragment)
S10786
C:Species: Bos primigenius taurus (cattle)
C>Date: 19-Mar-1997 #sequence_revision 21-Nov-1998 #text_change 21-Nov-1998
C:Accession: S10786
R:Strawich, E.; Glimcher, M.J.
Eur. J. Biochem. 191, 47-56, 1990
A:Title: Tooth 'enamelins' identified mainly as serum proteins. Major 'enamelin' is albu
A:Reference number: S10780; PMID:90336641; PMID:2379503
A:Accession: S10786
A:Molecule type: protein
A:Residues: 1-17 <STR>
C:Keywords: enamel; phosphoprotein

Query Match 25.4%; Score 29; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 HKKIP 11
|||
Db 10 HKKIP 14

RESULT 4

S43627
cytochrome-c oxidase (EC 1.9.3.1) chain Vb-H - trout (fragment)
C:Species: Salmo sp. (trout)
C>Date: 19-Mar-1997 #sequence_revision 01-Aug-1997 #text_change 02-Jul-1998
C:Accession: S43627
R:Freund, R.; Kadenbach, B.
Eur. J. Biochem. 221, 1111-1116, 1994
A:Title: Identification of tissue-specific isoforms for subunits Vb and VIIa of cytochr
A:Reference number: S43624; PMID:94237550; PMID:8181469
C:Accession: S43627
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-20 <PRE>
C:Keywords: electron transfer; membrane-associated complex; oxidoreductase; respiratory

Query Match 25.4%; Score 29; DB 2; Length 20;
Best Local Similarity 40.0%; Pred. No. 8.9e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 8 KPIPEEKREKLEQ 22
|||
Db 4 KPIPEDEGATGLEE 18

RESULT 5

S65385
cytochrome-c oxidase (EC 1.9.3.1) chain VIIa, hepatic - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 07-May-1999
C:Accession: S65385
R:Schaeffer, H.; Noack, H.; Halangsk, W.; Brandt, U.; von Jagow, G.
Eur. J. Biochem. 230, 235-241, 1995
A:Title: Cytochrome-c oxidase in developing rat heart. Enzymic properties and amino-term
A:Reference number: S65372; PMID:95524529; PMID:7601105
A:Accession: S65385
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-10 <SCH>
C:Keywords: oxidoreductase

Query Match 24.6%; Score 28; DB 2; Length 10;
Best Local Similarity 57.1%; Pred. No. 6.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 9 KIPEEK 15
|:|:|
Db 4 KVEPKOK 10

RESULT 6

164832
Ca2+-transporting ATPase (EC 3.6.3.8) - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 19-Apr-2002
C:Accession: 164832
R:Wu, X.
Am. J. Physiol. 264, 333-341, 1993
A:Title: Molecular cloning and quantification of mRNA and protein encoding Ca2+-ATPase 1
A:Reference number: 151892
A:Accession: 164832
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-8 <RES>
A:Cross-references: GB:M99223; NID:G203644; PIDN:AAA40992.1; PID:G203646
C:Genetics:
A:Gene: SERCALB
C:Keywords: hydrolase

Query Match 23.7%; Score 27; DB 2; Length 8;
Best Local Similarity 57.1%; Pred. No. 2.8e+05;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 11 PEEKRE 17
|||
Db 2 PEDERRK 8

RESULT 7

PA0041
plastocyanin-plastocyanin reductase (EC 1.10.99.1) - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 30-Jun-1992 #sequence_revision 06-Jan-1995 #text_change 03-Jun-2002
C:Accession: PA0041
R:Kamo, M.; Kawakami, T.; Miyatake, N.; Tsugita, A.
submitted to JIPID, July 1994
A:Description: Separation and characterization of Arabidopsis proteins by two-dimensiona
A:Reference number: PA0001
A:Accession: PA0041
A:Molecule type: protein
A:Residues: 1-15 <KAM>
A:Experimental source: leaf
C:Keywords: oxidoreductase

Query Match 23.7%; Score 27; DB 2; Length 15;
Best Local Similarity 50.0%; Pred. No. 1.3e+03;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 9 KIPEEK 16
|||
Db 8 RVPDMERK 15

RESULT 8

S13898
alkaline phosphatase (EC 3.1.3.1) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Aug-1997
C:Accession: S13898
R:Fujimori-Arai, Y.; Koyama, T.; Hirano, K.; Sakagishi, Y.; Komoda, T.
Arch. Biochem. Biophys. 284, 320-325, 1991
A:Title: Purification and partial characterization of intestinal-like alkaline phosphata
A:Reference number: S13898; PMID:9112827; PMID:1989515
A:Accession: S13898
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-16 <FUJ>
C:Keywords: phosphoric monoester hydrolase

Query Match 23.7%; Score 27; DB 2; Length 16;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 10 IPEEEK 15
DB 2 IPEEEK 7

RESULT 9

G83975
hypothetical protein BH2607 (imported) - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83975
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
C:Accession: G83975
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-17 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA806326.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2607

Query Match 23.7%; Score 27; DB 2; Length 17;
Best Local Similarity 50.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 8 KRIPEEKREKLE 21
DB 2 KEIPTKMKRIKDE 15

RESULT 10

A61220
epsilon receptor modulating protein (EC 3.4.21.-) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Mar-1996
C:Accession: A61220
R:Matsumita, S.; Katz, D.H.
Cell. Immunol. 137, 252-259, 1991
A:Title: The murine epsilon receptor modulating protein: a novel serine protease which m
A:Reference number: A61220; MUID:91356570; PMID:1679381
A:Accession: A61220
A:Molecule type: protein
A:Residues: 1-18 <MAT>
C:Comment: This serine proteinase from a T cell hybridoma does not reduce levels of CD23
C:Keywords: hydrolase; serine proteinase

Query Match 23.7%; Score 27; DB 2; Length 18;
Best Local Similarity 33.3%; Pred. No. 1.5e+03;
Matches 4; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19
DB 2 KPAPKKEK 13

RESULT 11

S11416
ribosomal protein L6, cytosolic [validated] - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 21-Jul-2000
C:Accession: S11416
R:Wittmann-Liebold, B.; Geiseler, A.W.; Lin, A.; Wool, I.G.
J. Supramol. Struct. 12, 425-433, 1979
A:Title: Sequence of the amino-terminal region of rat liver ribosomal proteins S4, S6, S
A:Reference number: S11413; MUID:80252792; PMID:398910

A:Accession: S11416
A:Molecule type: protein
A:Residues: 1-20 <MIT>
A>Note: the protein is designated as ribosomal protein L6
C:Keywords: protein biosynthesis; ribosome

Query Match 23.7%; Score 27; DB 2; Length 20;
Best Local Similarity 35.7%; Pred. No. 1.7e+03;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKRIPEEKREK 19
DB 1 AGEKAEKPKKEOK 14

RESULT 12

S36891
ribosomal protein - Mycobacterium bovis (fragment)
C:Species: Mycobacterium bovis
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Jan-1995
C:Accession: S36891
R:Ohara, N.; Kimura, M.; Higashi, Y.; Yamada, T.
FEBS Lett. 331, 9-14, 1993
A:Title: Isolation and amino acid sequence of the 30S ribosomal protein S19 from Mycobac
A:Reference number: S36887; MUID:94009653; PMID:8405418
C:Accession: S36891
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-15 <OHA>

Query Match 22.8%; Score 26; DB 2; Length 15;
Best Local Similarity 33.3%; Pred. No. 1.7e+03;
Matches 4; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19
DB 3 KNVPANSRRKAK 14

RESULT 13

E33178
76k heat shock protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 23-Mar-1993
C:Accession: E33178
R:Ward, L.D.; Hong, J.; Whitehead, R.H.; Simpson, R.J.
Electrophoresis 11, 883-891, 1990
A:Title: Development of a database of amino acid sequences for human colon carcinoma pro
A:Reference number: A33178; MUID:91176935; PMID:2079031
A:Accession: E33178
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-18 <MAR>

Query Match 22.8%; Score 26; DB 2; Length 18;
Best Local Similarity 57.1%; Pred. No. 2.1e+03;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EEKREE 18
DB 2 EEDKED 8

RESULT 14

PC7043
ubiquitin carboxyl-terminal hydrolase (EC 3.1.1.-) homolog - yeast (Pichia anomala) (fra
N:Alternate names: ALX1 protein
C:Species: Pichia anomala, Candida pelliculosa
C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 11-May-2000
C:Accession: PC7043
R:Sakajo, S.; Minagawa, N.; Yoshimoto, A.
Biosci. Biotechnol. Biochem. 63, 1889-1894, 1999
A:Title: Structure and regulated expression of a single copy alternative oxidase gene f

A:Reference number: JC7145; MUID:20101471; PMID:10635554
A:Accession: PC7043
A:Molecule type: DNA
A:Residues: 1-21 <SAK>
A:Cross-references: DDBJ:AB026726
C:Genetics:
A:Gene: alx1
C:Keywords: hydrolase

Query Match	22.8%	Score 26	DB 2	Length 21
Best Local Similarity	50.0%	Pred. No. 2.4e+03		
Matches	5	Conservative	3	Mismatches 2
				Indels 0
				Gaps 0

```
Qy      13 EKKREKLEQ 22
Db      12 EDKREGKEE 21
```

RESULT 15
F33064

ribosomal protein L18 - Haloferax mediterranei (fragment)
N/Alternate names: ribosomal protein HL13
C/Species: Haloferax mediterranei
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 18-Jun-1993
C/Accession: F33084
R/McDougal, J.
submitted to the Protein Sequence Database, June 1990
A/Reference number: A33084
A/Accession: F33084
A/Molecule type: protein
A/Residues: 1-22 <MCD>
A/Experimental source: strain DSM 1411
C/Superfamily: rat ribosomal protein L5
C/Keywords: protein biosynthesis; ribosome

Query Match	22.8%;	Score 26;	DB 2;	Length 22;
Best Local Similarity	30.8%;	Pred. No. 2.5e+03;		
Matches	4;	Conservative	4;	Mismatches 5;
				Indels .0;
				Gaps 0;

```
QY      9 KIPEEKREEKLE 21
          |::||:
Db      7 KVPMMMMRRREVRTD 19
```

Search completed: July 16, 2003, 08:14:42
Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:57:58 ; Search time 11 Seconds
(without alignments)
82.953 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFFAHKKRPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 1277

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	25.4	20	1	COXF_ONCMY
2	26.5	23.2	20	1	OXIA_CROAT
3	26	22.8	15	1	GR78_HORSE
4	26	22.8	22	1	RL18_HALME
5	25	21.9	15	1	MCA2_RHOOP
6	25	21.9	16	1	IBP4_PIG
7	25	21.9	20	1	COXB_THUOB
8	24.5	21.5	22	1	MOTI_CANRA
9	24	21.1	17	1	UP41_UPELN
10	24	21.1	18	1	LUXB_KRYAS
11	24	21.1	22	1	MOTI_CHICK
12	24	21.1	22	1	NHO7_SOLRU
13	23	20.2	13	1	IDHP_RAT
14	23	20.2	15	1	UNO4_PINS
15	23	20.2	20	1	PSAF_MAIZE
16	23	20.2	22	1	FUC1_RAT
17	23	20.2	22	1	FUC2_RAT
18	23	20.2	22	1	RL41_METJA
19	22	19.3	10	1	COXA_ONCMY
20	22	19.3	10	1	PAP1_PARMA
21	22	19.3	14	1	RS19_PRUAP
22	22	19.3	20	1	TL14_SPIOL
23	22	19.3	21	1	FIIB_CEREL
24	21	18.4	12	1	HS9A_RAT
25	21	18.4	15	1	COXJ_THUOB
26	21	18.4	17	1	UN15_CLOFA
27	21	18.4	19	1	PSBM_SYNUU
28	21	18.4	20	1	COXA_THUOB
29	20	17.5	10	1	SRTD_ATREN
30	20	17.5	10	1	COXK_ONCMY
31	20	17.5	10	1	TKU2_UREIN
32	20	17.5	10	1	URE3_MORNO
33	20	17.5	12	1	RS19_CLVEP

34	20	17.5	12	1	RS19_TOBBP	Q66251 tomato big
35	20	17.5	13	1	E121_LITRU	P82097 litorea rub
36	20	17.5	15	1	E122_LITRU	P82098 litorea rub
37	20	17.5	15	1	COXI_THUOB	P80978 thunnus obe
38	20	17.5	20	1	GMS2_ASCSU	P48429 ascaris suu
39	20	17.5	21	1	BRX_ATRBI	P80163 atracaspis
40	20	17.5	21	1	CKXT_CONTU	P17684 conus tulip
41	20	17.5	21	1	FIIB_ANTAM	P14465 antilocapra
42	20	17.5	21	1	RL5_HALME	P50557 halobacteri
43	20	17.5	22	1	RL5_HALVO	P50559 halobacteri
44	19	16.7	5	1	RE32_LITRU	P82073 litorea rub
45	19	16.7	10	1	BPP2_BOTIN	P30422 boctrops in

ALIGNMENTS

RESULT 1
COXF_ONCMY STANDARD; PRT; 20 AA.
AC P80329;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cytochrome c oxidase polypeptide Vb-heart (EC 1.9.3.1) (Fragment).
OS Oncoerynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Proactanthopterygii; Salmoniformes; Salmonidae; Oncoerynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE.
RC TISSUE=Heart;
RX MEDLINE=94237150; PubMed=8181469;
RA Freund R.; Kadenbach B.;
RT "Identification of tissue-specific isoforms for subunits Vb and Villa
of cytochrome c oxidase isolated from rainbow trout.";
RL Eur. J. Biochem. 221:1111-1116(1994).
CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
MITOCHONDRIAL ELECTRON TRANSPORT.
CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
c + 2 H(2)O.
CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VB FAMILY.
DR InterPro: IPR002124; COX5B.
DR PROSITE; PS00846; COX5B; PARTIAL.
KW Oxidoreductase; Inner membrane; Mitochondrion.
FT NON TER 20
SQ SEQUENCE 20 AA; 2184 MW; 6A8B0C33E92BD7 CRC64;
Query March 25.4%; Score 29; DB 1; Length 20;
Best Local Similarity 40.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 8 KKIPEEKREKLEQ 22
4 KGIPTDEQATGLEE 18
RESULT 2
OXIA_CROAT STANDARD; PRT; 20 AA.
AC P56742;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE L-amino acid oxidase (EC 1.4.3.2) (LAO) (Apoxi I) (Fragment).
OS Crotalaria atrox (Western diamondback rattlesnake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Chordata; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Crotalus.
OX NCBI_TaxID=8730;

```

RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RX MEDLINE=97238898; PubMed=9083096;
RA Torii S., Naito M., Tsuno T.;
RT "Apoxin I, a novel apoptosis-inducing factor with L-amino acid oxidase
RL activity purified from Western diamondback rattlesnake venom.";
RJ J. Biol. Chem. 272:9539-9542(1997).
CC -1- FUNCTION: CATALYZES AN OXIDATIVE DEMINATION OF PERDOMINANTLY
CC HYDROPHOBIC AND AROMATIC L-AMINO ACIDS. HAS AN ANTIBACTERIAL
CC EFFECT AND AN ABILITY TO INDUCE APOPTOSIS. THE H(2)O(2) PRODUCED
CC BY L-AMINO ACID OXIDATION IS INVOLVED IN THE APOXIN-I INDUCED
CC APOPTOSIS AND HEMORRHAGE CAUSED BY THE VENOM.
CC -1- CATALYTIC ACTIVITY: An L-amino acid + H(2)O + O(2) = a 2-oxo acid
CC + NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE FLAVIN MONOAMINE OXIDASE FAMILY.
CC STRONG. TO MOUSE FIG-1.
KM Oxidoreductase; Flavoprotein; FAD; Venom; Apoptosis; Antibiotic.
FT NON TER 20 20
SQ SEQUENCE 20 AA; 2521 MW; 95E973C6C86588E7 CRC64;

Query Match 23.2%; Score 26.5; DB 1; Length 20;
Best Local Similarity 44.4%; Pred. No. 8e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

QY 6 AHKIP-EKKREKLEQ 22
DB 1 AHDRLPLEXPRETDYEE 18

RESULT 3
GR78_HORSE
ID GR78_HORSE STANDARD; PRT; 15 AA.
AC P16392;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 78 kDa glucose-regulated protein (GRP 78) (Immunoglobulin heavy chain
DE binding protein) (BIP) (Fragment).
GN HSPA5 OR GRP78.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Perissodactyla; Equidae; Equus.
CX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RX MEDLINE=90147817; PubMed=1689156;
RA Obias B., Boyd N.D., Lubber-Narod J., Reyes V.E., Leeman S.E.;
RT "Isolation and identification of a polypeptide in the Hsp 70 family
RT that binds substance P.";
RL Biochem. Biophys. Res. Commun. 166:978-983(1990).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR InterPro; IPR001023; Hsp70.
DR PROSITE; PS00297; HSP70_1; PARTIAL.
DR PROSITE; PS00329; HSP70_2; PARTIAL.
DR PROSITE; PS01036; HSP70_3; PARTIAL.
KM ATP-binding; Endoplasmic reticulum.
FT NON TER 15 15
SQ SEQUENCE 15 AA; 1647 MW; 81119D21D0EC26DB CRC64;

Query Match 22.8%; Score 26; DB 1; Length 15;
Best Local Similarity 57.1%; Pred. No. 7e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EKKREE 18
DB 2 EEDKED 8

```

```

RESULT 4
RL18_HALME
ID RL18_HALME STANDARD; PRT; 22 AA.
AC P50561;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L18P (HmeL18) (Fragment).
GN RPL18P.
OS Halobacterium mediterranei (Haloflex mediterranei).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloflex.
CX NCBI_TaxID=2252;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 1411;
RX MEDLINE=94229075; PubMed=8174557;
RA McDougall J., Wiltmann-Liebold B.;
RT "Comparative analysis of the protein components from 5S rRNA protein
RL complexes of halophilic archaeobacteria.";
RT Eur. J. Biochem. 221:779-785(1994).
CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.
CC PIR; F33084; F33084.
KM Ribosomal protein.
FT NON TER 22 22
SQ SEQUENCE 22 AA; 2773 MW; FE1F607FCAE9876D CRC64;

Query Match 22.8%; Score 26; DB 1; Length 22;
Best Local Similarity 30.8%; Pred. No. 1e+03;
Matches 4; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 KIPPEKREKLE 21
DB 7 KVPKRRRREVRTD 19

RESULT 5
MCA2_RHOOP
ID MCA2_RHOOP STANDARD; PRT; 15 AA.
AC P56870;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative malylacetate reductase II (EC 1.3.1.32) (Fragment).
OS Rhodococcus opacus (Nocardia opaca).
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
CX NCBI_TaxID=37919;
RN [1]
RP SEQUENCE.
RC STRAIN=ICP;
RX MEDLINE=98324954; PubMed=9657989;
RA Seibert V., Koubatova E.M., Golovleva L.A., Schloemann M.;
RT "Characterization of a malylacetate reductase encoding region from
RT Rhodococcus opacus ICP.";
RL J. Bacteriol. 180:3503-3508(1998).
CC -1- CATALYTIC ACTIVITY: 3-oxoadipate + NAD(P)(+) = 2-malylacetate +
CC NAD(P)H.
CC -1- PATHWAY: 3-CHLOROCATECHOL DEGRADATION (BETA-KETOADIPATE PATHWAY).
CC THIS PATHWAY SERVES A VITAL ROLE IN THE BIODEGRADATION OF TOXIC
CC AROMATIC COMPOUNDS INTRODUCED IN THE ENVIRONMENT BOTH AS NATURAL
CC PRODUCTS AND AS INDUSTRIAL EFFLUENT.
CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
CC FAMILY.
DR InterPro; IPR001670; Fe-ADH.
DR PROSITE; PS00913; ADH_IRON_1; PARTIAL.
DR PROSITE; PS00060; ADH_IRON_2; PARTIAL.
KM Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
FT NON TER 15 15
SQ SEQUENCE 15 AA; 1884 MW; 58DA9DD038F025E CRC64;

```

Query Match 21.9%; Score 25; DB 1; Length 15;
 Best Local Similarity 37.5%; Pred. No. 9.6e+02;
 Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 FAKKPEEK 12
 DB 3 FEHENDPO 10

RESULT 6

ID IBP4_PIG STANDARD; PRT; 16 AA.
 AC P24854;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Insulin-like growth factor binding protein 4 (IGFBP-4) (IBP-4)
 DE (IGF-binding protein 4) (Fragment).
 GN IGFBP4.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=96823;
 RN [1]

RP SEQUENCE.
 RX MEDLINE=92109718; PubMed=1722398;
 RA Coleman M.E., Pan Y.-C.E., Eberhart T.D.;
 RT "Identification and NH₂-terminal amino acid sequence of three
 insulin-like growth factor-binding proteins in porcine serum.";
 RL Biochem. Biophys. Res. Commun. 181:1131-1136(1991).
 CC -1- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
 AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
 PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
 INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
 PROTEIN FAMILY.
 CC PIR, JH0517; JH0517.
 DR InterPro; IPR000867; Insl_gro_fac.pr.
 DR InterPro; IPR000716; Thyroglobulin_1.
 DR PROSITE; PS00222; IGF_BINDING, PARTIAL.
 DR PROSITE; PS00484; THYROGLOBULIN_1; PARTIAL.
 KM Growth factor binding.
 FT NON TER 16
 SQ SEQUENCE 16 AA; 1799 MW; 40988840096655E2 CRC64;

Query Match 21.9%; Score 25; DB 1; Length 16;
 Best Local Similarity 55.6%; Pred. No. 1e+03;
 Matches 5; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 7 HKKPEEEK 15
 DB 5 HAKPPSEK 13

RESULT 7

ID COXB_THIOB STANDARD; PRT; 20 AA.
 AC P80974;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Cytochrome c oxidase polypeptide Vb (EC 1.9.3.1) (Fragment).
 OS Thynnus obesus (Bigeye tuna).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Scombroidei;
 OC Scombridae; Thunnus.
 OX NCBI_TaxID=8241;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Heart, and Liver;

RX MEDLINE=97454291; PubMed=9310366;
 RA Arnold S., Lee I., Kim M., Song E., Linder D., Lottspeich F.,
 RA Kadenbach B.;
 RT "The subunit structure of cytochrome-c oxidase from tuna heart and
 liver.";
 RL Eur. J. Biochem. 248:99-103(1997).
 CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE NUCLEAR-CODED POLYPEPTIDE
 CHAINS OF CYTOCHROME C OXIDASE, THE TERMINAL OXIDASE IN
 MITOCHONDRIAL ELECTRON TRANSPORT.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VB FAMILY.
 DR InterPro; IPR002124; COX5B.
 DR PROSITE; PS00848; COX5B; PARTIAL.
 KM Oxidoreductase; Inner membrane; Mitochondrion.
 FT NON TER 20
 SQ SEQUENCE 20 AA; 2158 MW; 76F1473E1F392BD7 CRC64;

Query Match 21.9%; Score 25; DB 1; Length 20;
 Best Local Similarity 33.3%; Pred. No. 1.3e+03;
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 8 KKIPEEKREKLPQ 22
 DB 4 KGIPTDEQATGLER 18

RESULT 8

ID MOT1_CANFA STANDARD; PRT; 22 AA.
 AC P19863;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Motilin.
 GN Motilin.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Intestine;
 RX MEDLINE=83195948; PubMed=684463;
 RA Poltias P., Reeve J.R., Hunkapiller M.W., Hood L.E., Walsh J.H.;
 RT "Purification and characterization of canine intestinal motilin.";
 RL Regul. Pept. 5:197-208(1983).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE REGULATION OF
 INTERDIGESTIVE GASTROINTESTINAL MOTILITY AND INDIRECTLY CAUSES
 RHYTHMIC CONTRACTION OF DUODENAL AND COLONIC SMOOTH MUSCLE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE MOTILIN FAMILY.
 DR PIR, S00189; S00189.
 DR PIR, A60313; A60313.
 KM Hormone.
 FT UNSURE 1
 SQ SEQUENCE 22 AA; 2685 MW; 4BECB840ABE0639F CRC64;

Query Match 21.5%; Score 24.5; DB 1; Length 22;
 Best Local Similarity 37.5%; Pred. No. 1.6e+03;
 Matches 6; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

OY 5 FAH--KKIPEEKRE 17
 DB 5 FTTHSELOKIRKERNK 20

RESULT 9

ID UP41_UPEIN STANDARD; PRT; 17 AA.
 AC P82035;
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Uperin 4.1.
 OS Uperoleia inundata (Floodplain toadlet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Myobatrachidae;
 OC Uperoleia.
 NC NCB1_Taxid=104953;
 RN [1]
 RP SEQUENCE, AND MASS SPECTROMETRY.
 RC TISSUE=Skin secretion;
 RA Bradford A.M., Rattery M.J., Bowie J.H., Tyler M.J., Wallace J.C.,
 Adams G.W., Severini C.;
 RT "Novel uperin peptides from the dorsal glands of the Australian
 floodplain toadlet Uperoleia inundata.";
 AU Aust. J. Chem. 49:475-484 (1996)
 CC -1- FUNCTION: SHOWS A MEDIUM ANTIBACTERIAL ACTIVITY AGAINST
 L. MENSETERIODES.
 CC -1- MASS SPECTROMETRY: MW=1724; METHOD=FAB.
 KW Amphibian skin; Antibiotic; Amidation.
 FT MOD RES 17 17
 SO SEQUENCE 17 AA; 1726 MW; 62E868714BC54D7B CRC64;
 QY
 Query Match 21.1%; Score 24; DB 1; Length 17;
 Best Local Similarity 44.4%; Pred. No. 1.5e+03;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 DB 2 LGFPAHKKI 10
 2 VGSFHKV 10
 RESULT 10
 LUXB_KRYAS STANDARD; PRT; 18 AA.
 AC P18300;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Alkanal monooxygenase beta chain (EC 1.14.14.3) (Bacterial luciferase
 beta chain) (Fragment).
 GN LUXB.
 OS Kryptophanon alfredi symbiont.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
 OC light emitting symbionts of fish.
 NC NCB1_Taxid=28177;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91076680; PubMed=2256783;
 RA Haygood M.G.;
 RT "Relationship of the luminous bacterial symbiont of the Caribbean
 flashlight fish, Kryptophanon alfredi (family Anamolopidae) to
 other luminous bacteria based on bacterial luciferase (luxA) genes.";
 RL Arch. Microbiol. 154:496-503 (1990).
 CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
 SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
 REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: RCHO + FMN(2) + O(2) = RCOOH + FMN + H(2)O +
 light.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isd-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, M36597; AA091214.1; -
 DR PIR, B44995; B44995.
 DR InterPro, IPR002103; Bac_Luciferase.
 DR PROSITE, PS00494; BACTERIAL_LUCIFERASE; PARTIAL.

KW Photoprotein; Luminescence; Oxidoreductase; Monooxygenase;
 KW Flavoprotein; FMN.
 FT NON TER 18
 SQ SEQUENCE 18 AA; 2153 MW; 8CB3B695CCB2E7A CRC64;
 QY
 Query Match 21.1%; Score 24; DB 1; Length 18;
 Best Local Similarity 29.4%; Pred. No. 1.6e+03;
 Matches 5; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
 DB 1 KLGFPAHKKIPEEKRE 17
 2 KFGLEFQNFLENSOSE 18
 RESULT 11
 MOTI_CHICK STANDARD; PRT; 22 AA.
 ID MOTI_CHICK
 AC Q9PRP6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Motilin.
 GN Motilin.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCB1_Taxid=9031;
 RN [1]
 RP SEQUENCE.
 RC TISSUE=Small intestine;
 RX MEDLINE=96255357; PubMed=8801522;
 RA De Clercq P., Depoortere I., Maclellan M., Vandermeers A.,
 Vandermeers-Piret M.C., Peeters T.L.;
 RT "Isolation, sequence, and bioactivity of chicken motilin.";
 RL Peptides 17:203-208 (1996).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE REGULATION OF
 INTERDIGESTIVE GASTROINTESTINAL MOTILITY AND INDIRECTLY CAUSES
 RHYTHMIC CONTRACTION OF DUODENAL AND COLONIC SMOOTH MUSCLE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE MOTILIN FAMILY.
 KW Hormone.
 SQ SEQUENCE 22 AA; 2686 MW; 6FD6A79B7EB02680 CRC64;
 QY
 Query Match 21.1%; Score 24; DB 1; Length 22;
 Best Local Similarity 26.3%; Pred. No. 1.9e+03;
 Matches 5; Conservative 4; Mismatches 10; Indels 0; Gaps 0;
 DB 4 FFAHKKIPEEKREKLEQ 22
 4 FFTQSDIQKQEKERNKQ 22
 RESULT 12
 NUD7 SOLTU STANDARD; PRT; 22 AA.
 ID NUD7 SOLTU
 AC P80730;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE NADH-ubiquinone oxidoreductase 16 kDa subunit (EC 1.6.5.3)
 DE (EC 1.6.99.3) (Complex I-16KD) (CI-16KD) (Fragment).
 OS Solanum tuberosum (Potato).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 NC NCB1_Taxid=4113;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=cv. Bintje; TISSUE=Tuber;
 RA Herz U., Gschmann L.;
 RT Submitted (DEC-1996) to the SWISS-PROT data bank.
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY


```

CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -1- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MATRIX SIDE OF THE MITOCHONDRIAL INNER
CC MEMBRANE.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
KM Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON TER 22
SQ SEQUENCE 22 AA; 2435 MW; D4D64C57673C2613 CRC64;

Query Match 21.1%; Score 24; DB 1; Length 22;
Best Local Similarity 62.5%; Pred. No. 1.9e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 8 KKIPEEK 15
DB 4 KVLSEEEK 11

RESULT 13
IDHP RAT STANDARD; PRT; 13 AA.
AC P56574;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Isocitrate dehydrogenase [NADP], mitochondrial (EC 1.1.1.42)
DE (Oxalosuccinate decarboxylase) (IDH) (NADP+-specific ICDH) (IDP) (ICD-
DE M) (Fragment).
GN IDH2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RC STRAIN=Wiistar; TISSUE=Heart;
RA Li X.-P., Pleisner K.-P., Scheler C., Regitz-Zagrosek V., Salikov J.,
RA Jungblut P.R.;
RL Submitted (SEP-1998) to the SWISS-PROT data bank.
CC -1- FUNCTION: PLAYS A ROLE IN INTERMEDIARY METABOLISM AND ENERGY
CC PRODUCTION. IT MAY TIGHTLY ASSOCIATE OR INTERACT WITH THE PYRUVATE
CC DEHYDROGENASE COMPLEX (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Isocitrate + NADP(+) = 2-oxoglutarate + CO(2)
CC + NADH.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC (SPOT P8) IS: 9.0, ITS MW IS: 42 kDa.
CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC DEHYDROGENASES FAMILY.
CC InterPro; IPR001804; Iscdh.
DR PROSITE; PS00470; IDH_IDMH; PARTIAL.
KM Oxidoreductase; NADP; Glyoxylate bypass; Tricarboxylic acid cycle;
KM Mitochondrion.
FT NON TER 13
SQ SEQUENCE 13 AA; 1526 MW; 5FB81031723B02C3 CRC64;

Query Match 20.2%; Score 23; DB 1; Length 13;
Best Local Similarity 50.0%; Pred. No. 1.5e+03;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 AHKKIPEEK 15
DB 1 AEKRIKVEK 10

RESULT 14
UN04_PINPS STANDARD; PRT; 15 AA.
AC P81673;
DT 15-JUL-1999 (Rel. 38, Created)

```

```

DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Unknown protein from 2D-page of needles (N143) (Fragment).
OS Pinus pinaster (Maritime pine).
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=71647;
RN [1]
RP SEQUENCE.
RC TISSUE=Needle;
RX MEDLINE=99274088; PubMed=10344291;
RA Costa P., Pionneau C., Bauw G., Dubos C., Bahman N., Kremer A.,
RA Frigerio J.-M., Plomion C.;
RT "Separation and characterization of needle and xylem maritime pine
RT proteins.";
RL Electrophoresis 20:1098-1108(1999).
CC -1- INDUCTION: BY WATER STRESS.
CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 6.2, ITS MW IS: 21 kDa.
FT NON TER 1
FT NON TER 15
SQ SEQUENCE 15 AA; 1489 MW; CE4D85E9308227A CRC64;

Query Match 20.2%; Score 23; DB 1; Length 15;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 KKIPEE 13
DB 1 KKIIEE 6

RESULT 15
PSAF MAIZE STANDARD; PRT; 20 AA.
ID P13193;
AC P13193;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Photosystem I reaction centre subunit III (Light-harvesting complex I
DE 17 kDa protein) (Pst-P) (Fragment).
GN PSAP.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE.
RC STRAIN=cv. N273;
RX MEDLINE=90033290; PubMed=2680596;
RA Anandan S., Vainstein A., Thorner J.P.;
RT "Correlation of some published amino acid sequences for photosystem I
RT polypeptides to a 17 kDa LHCI pigment-protein and to subunits III and
RT IV of the core complex.";
RL FEBS Lett. 256:150-154(1989).
CC -1- FUNCTION: PROBABLY PARTICIPATES IN EFFICIENCY OF ELECTRON TRANSFER
CC FROM PLASTOCYANIN TO P700 (OR CYTOCHROME C553 IN ALGAE AND
CC CYANOBACTERIA). THIS PLASTOCYANIN-DOCKING PROTEIN CONTRIBUTES
CC TO THE SPECIFIC ASSOCIATION OF PLASTOCYANIN TO PS I.
CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH LUMENAL SIDE OF THE
CC THYLAKOID MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE PSAF FAMILY.
CC PIR; S06150; S06150.
DR MaizeDB; 69342;
KM Photosynthesis; Photosystem I; Chloroplast; Thylakoid; Membrane.
FT NON TER 20
SQ SEQUENCE 20 AA; 2110 MW; F0499F98F6188997 CRC64;

Query Match 20.2%; Score 23; DB 1; Length 20;
Best Local Similarity 41.7%; Pred. No. 2.4e+03;
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

Oy 11 PEEKREKLEQ 22
|:|:|:
Db 8 PKEKAFAXEK 19

Search completed: July 16, 2003, 08:13:46
Job time : 13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:07:08 ; Search time 27 Seconds
(without alignments)
167.890 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 7500

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	27.2	19	13	Q91983
2	30	26.3	13	10	Q9380
3	30	26.3	20	4	Q16070
4	30	26.3	22	4	Q9R503
5	29.5	25.9	20	6	Q9R7Y7
6	28	24.6	22	4	Q9UC03
7	27.5	24.1	15	2	Q9R5Z5
8	27	23.7	16	4	Q9UC12
9	27	23.7	17	16	Q9K5N8
10	27	23.7	18	6	Q9R7G8
11	27	23.7	20	4	Q15969
12	27	23.7	22	5	Q44710
13	26	22.8	14	11	Q9J3U5
14	26	22.8	15	2	Q9R545
15	26	22.8	15	4	Q9UBK0
16	26	22.8	15	10	P83137

17	26	22.8	19	4	Q9UG7	Q9UG7 homo sapien
18	25	21.9	9	12	Q9Z766	Q9Z766 canine dist
19	25	21.9	16	4	Q9NZH9	Q9NZH9 homo sapien
20	25	21.9	16	6	Q9TR09	Q9TR09 bos taurus
21	25	21.9	16	10	P83142	P83142 malva parvi
22	25	21.9	18	13	Q9PRR7	Q9PRR7 gallus galli
23	25	21.9	19	12	Q65747	Q65747 blueongue
24	25	21.9	20	2	Q9R4Q0	Q9R4Q0 pseudomonas
25	25	21.9	20	4	Q75318	Q75318 homo sapien
26	25	21.9	20	4	Q76086	Q76086 homo sapien
27	25	21.9	20	4	Q9UC8	Q9UC8 homo sapien
28	25	21.9	21	2	Q9R4T3	Q9R4T3 mycobacteri
29	25	21.9	21	10	Q41559	Q41559 triticum ae
30	24.5	21.5	15	10	Q41526	Q41526 triticum ae
31	24	21.1	13	2	Q55234	Q55234 synechocyst
32	24	21.1	15	6	Q9N0F7	Q9N0F7 canis famli
33	24	21.1	16	4	Q9UD21	Q9UD21 homo sapien
34	24	21.1	16	6	Q9N0F6	Q9N0F6 canis famli
35	24	21.1	16	9	Q38671	Q38671 bacterioph
36	24	21.1	17	4	Q9UC6	Q9UC6 homo sapien
37	24	21.1	17	6	Q9TR21	Q9TR21 sus scrofa
38	24	21.1	17	11	Q9QVS7	Q9QVS7 mus sp. lac
39	24	21.1	18	15	Q73573	Q73573 human immun
40	24	21.1	18	15	Q73583	Q73583 human immun
41	24	21.1	18	15	Q73587	Q73587 human immun
42	24	21.1	18	15	Q73603	Q73603 human immun
43	24	21.1	18	15	Q73605	Q73605 human immun
44	24	21.1	18	15	Q73607	Q73607 human immun
45	24	21.1	18	15	Q73613	Q73613 human immun

ALIGNMENTS

RESULT 1

ID	Q91983	PRELIMINARY;	PRT;	19 AA.
AC	Q91983;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)			
DE	Myosin heavy chain (Fragment).			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RP	SEQUENCE FROM N.A.			
RP	McCarthy J.G., Heywood S.M.,			
RA	"A long polypyrimidine/polyurine tract induces an altered DNA conformation on the 3' coding region of the adjacent myosin heavy chain gene."			
RL	Nucleic Acids Res. 15:8069-8085(1987).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	McCarthy J.G.,			
RL	Submitted (FEB-1989) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=89210285; PubMed=2707122;			
RA	Lagrutta A.A., McCarthy J.G., Scherzinger C.A., Heywood S.M.,			
RT	"Identification and developmental expression of a novel embryonic myosin heavy-chain gene in chicken."			
RT	DNA 8:39-50(1989).			
DR	EMBL; X06251; CAA29593.1; -			
KW	Myosin.			
FT	NON_TER			
SQ	SEQUENCE 19 AA; 2387 MW; 2DD5C6816CD4D6EC CRC64;			
Query Match		27.2%;	Score 31;	DB 13; Length 19;

Best Local Similarity 70.0%; Pred. No. 1e+03;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 FAHKKIPPEE 14
| | | | |
DB 10 FHSKKIEEE 19

RESULT 2

Q39380 PRELIMINARY; PRT; 13 AA.

AC Q39380; PRELIMINARY; PRT; 13 AA.
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE IFA binding protein (Sp10) (Fragment).
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside II; Brassicales; Brassicaceae; Brassica.
OX NCBI_Taxid=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DOK; TISSUE=CURD SURFACE;
RA Willis G.;
RT "An investigation of nuclear lamin homologues in plants: an apparently
RT non-intermediate filament sequence that bind a polyclonal anti-lamin
RT antiserum.";
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X57678; CAA66268.1; -.
FT NON_TER 1
SQ SEQUENCE 13 AA; 1413 MW; DID4EA3926B42772 CRC64;

Query Match 26.3%; Score 30; DB 10; Length 13;
Best Local Similarity 62.5%; Pred. No. 9.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 AHKKIPPEE 13
| | | | |
DB 2 AEKKVPKE 9

RESULT 3
Q16070 PRELIMINARY; PRT; 20 AA.
AC Q16070;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)
DE Neurofilament heavy subunit (Fragment).
GN NEFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040777; PubMed=8224877;
RA Figlewicz D.A., Rouleau G.A., Krizus A., Julien J.P.;
RT "Polymorphism in the multi-phosphorylation domain of the human
RL Gene 132:297-300(1993).
DR EMBL; S66488; AAB28609.1; -.
FT NON_TER 1
SQ SEQUENCE 20 AA; 2198 MW; E9A0975B41FD8082 CRC64;

Query Match 26.3%; Score 30; DB 4; Length 20;
Best Local Similarity 53.8%; Pred. No. 1.5e+03;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 9 KIPEEKREKLE 21
| | | | |
DB 6 KSPERAKSPEKEE 18

RESULT 4

Q9RSU3 PRELIMINARY; PRT; 22 AA.

AC Q9RSU3;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE 2-amino-benzoate-CoA LIGASE=E2 (Fragment).
OS Pseudomonas.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.
OX NCBI_Taxid=286;
RN [1]
RP SEQUENCE.
RX MEDLINE=91358327; PubMed=1885526;
RA Altmanshmidt U., Oswald B., Fuchs G.;
RT "Purification and characterization of benzoate-coenzyme A ligase and
RT 2-amino-benzoate-coenzyme A ligases from a dentrifying Pseudomonas
RT sp.";
RL J. Bacteriol. 173:5494-5501(1991).
FT NON_TER 1
FT NON_TER 22
SQ SEQUENCE 22 AA; 2526 MW; 5D37FA76363F8792 CRC64;

Query Match 26.3%; Score 30; DB 2; Length 22;
Best Local Similarity 37.5%; Pred. No. 1.6e+03;
Matches 6; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 5 FAHKKIPPEEKREKL 20
| | | | |
DB 7 FARDBXPTPEQOTESTL 22

RESULT 5

Q9TRY7 PRELIMINARY; PRT; 20 AA.

AC Q9TRY7;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Insulin-like growth factor-binding protein-5, IGFBP-5 (Fragment).
OS Sus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_Taxid=9826;
RN [1]
RP SEQUENCE.
RX MEDLINE=92049376; PubMed=1719383;
RA Shimasaki S., Gao L., Shimomura M., Ling N.;
RT "Isolation and molecular cloning of insulin-like growth factor-binding
RL protein-6.";
RL Mol. Endocrinol. 5:938-948(1991).
FT NON_TER 1
FT NON_TER 20
SQ SEQUENCE 20 AA; 2185 MW; F77C34A8FA94EF9 CRC64;

Query Match 25.9%; Score 29.5; DB 6; Length 20;
Best Local Similarity 50.0%; Pred. No. 1.7e+03;
Matches 7; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 2 LGFFAHKKIPPEEK 15
| | | | |
DB 1 LGSFVHKE-PYDERK 13

RESULT 6

Q9UCO3 PRELIMINARY; PRT; 22 AA.

AC Q9UCO3;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)

DE KU antigen 73 kDa protein fraction 31 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Carnivora; Homiidae; Homo.
RN NCB1_TaxID=9606;
[1]
RP SEQUENCE.
RX MEDLINE=92165607; PubMed=1537839;
RA Medychowski A., Henzel W., Huston L., Paslidis N., Ellerson D.,
McRae M., Seong D., Howard O.M., Deiseroth A.;
RT "Identification of proteins binding to interferon-inducible
transcriptional enhancers in hematopoietic cells."
RL J. Biol. Chem. 267:4533-4540 (1992).
FT NON_TER 1
FT TER 22
SQ SEQUENCE 22 AA; 2488 MW; 24E2B59F08AF2996 CRC64;

Query Match 24.6%; Score 28; DB 4; Length 22;
Best Local Similarity 66.7%; Pred. No. 3.1e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 13 EEKREKLE 21
DB 8 EEDEENLE 16

RESULT 7
O9R5Z5 PRELIMINARY; PRT; 15 AA.
AC O9R5Z5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Streptolysin O (Fragment).
OS Streptococcus equisimilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
RN NCB1_TaxID=119602;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93273535; PubMed=8500915;
RA Gerlach D., Kohler W., Gunther E., Mann K.;
RT "Purification and characterization of streptolysin O secreted by
Streptococcus equisimilis (group C)."
RL Infect. Immun. 61:2727-2731 (1993).
SQ SEQUENCE 15 AA; 1716 MW; 7C02547F72FF85BE CRC64;

Query Match 24.1%; Score 27.5; DB 2; Length 15;
Best Local Similarity 46.7%; Pred. No. 2.5e+03;
Matches 7; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

OY 6 AHKRIIP-EEKREK 19
DB 1 APKEMPLESAEKREK 15

RESULT 8
O9UC12 PRELIMINARY; PRT; 16 AA.
AC O9UC12;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Tropomyosin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Carnivora; Homiidae; Homo.
RN NCB1_TaxID=9606;
[1]
RP SEQUENCE.
RX MEDLINE=93195352; PubMed=8450225;
RA Das K.M., Dasgupta A., Mandal A., Geng X.;
RT "Autoimmunity to cytoskeletal protein tropomyosin. A clue to the

RT pathogenetic mechanism for ulcerative colitis.";
RL J. Immunol. 150:2487-2493 (1993).
DR InterPro: IPR000533; Tropomyosin.
DR Pfam: PF00261; Tropomyosin.1.
SQ SEQUENCE 16 AA; 1960 MW; 7A26C49A1E0A335E CRC64;

Query Match 23.7%; Score 27; DB 4; Length 16;
Best Local Similarity 62.5%; Pred. No. 3.2e+03;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 14 EKEKREKLE 21
DB 2 QKDEKXME 9

RESULT 9
O9K9N8 PRELIMINARY; PRT; 17 AA.
AC O9K9N8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein BH2607.
GN BH2607.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
RN NCB1_TaxID=86665;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takemi H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
Fuji F., Hirama C., Nakamura Y., Ogawara N., Kuhara S.,
Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331 (2000).
DR EMBL: AP001516; BAB06326.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 17 AA; 2086 MW; 2238BAPDE271681F CRC64;

Query Match 23.7%; Score 27; DB 16; Length 17;
Best Local Similarity 50.0%; Pred. No. 3.4e+03;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 8 KKIPEEKREKLE 21
DB 2 KEIPTKMKMIKDE 15

RESULT 10
O9TRG8 PRELIMINARY; PRT; 18 AA.
AC O9TRG8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Lactoferrin (Fragment).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN NCB1_TaxID=9913;
[1]
RP SEQUENCE.
RX MEDLINE=93253156; PubMed=8486845;
RA Shimazaki K., Tanaka T., Kon H., Oota K., Kawaguchi A., Maki Y.,
Sato T.;
RL J. Dairy Sci. 76:946-955 (1993).
SQ SEQUENCE 18 AA; 2138 MW; 809196895D52D7CE CRC64;

Query Match 23.7%; Score 27; DB 6; Length 18;

Best Local Similarity 50.0%; Pred. No. 3.6e+03;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 PEEKREE 18
| | | | |
Db 11 PEOKKXQ 18

RESULT 11

015969 PRELIMINARY; PRT; 20 AA.
AC 015969;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE Apolipoprotein A-IV (Fragment).
GN APOA-IV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed:1487136;
RX MEDLINE=93138374; PubMed:1487136;
RA Kamboh M.I., Williams E.R., Law J.C., Aston C.E., Bunker C.H.,
Ferrell R.E., Pollitzer W.S.;
RT "Molecular basis of a unique African variant (A-IV 5) of human
apolipoprotein A-IV and its significance in lipid metabolism.";
RL Genet. Epidemiol. 9:379-388(1992).
DR EMBL; S53635; AAB50137.1; -.
KW Lipoprotein.
FT NON_TER 1
SQ SEQUENCE 20 AA; 2557 MW; 03B96D8A967D7DB CRC64;

Query Match 23.7%; Score 27; DB 4; Length 20;
Best Local Similarity 36.4%; Pred. No. 3.9e+03;
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 12 EEEKREKLEQ 22
| | | | |
Db 4 QEOQOEQQOEQ 14

RESULT 12

044710 PRELIMINARY; PRT; 22 AA.
AC 044710;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Tropomyosin 1 isoform B (Fragment).
GN TMI.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7237;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GOLDENDALE 3;
RA Hamlin M.T.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF039274; AAB96670.1; -.
DR EMBL; AF039273; AAB96670.1; JOINED.
DR FlyBase; FBgn0025413; Dpse\Tmi.
DR InterPro; IPR000533; Tropomyosin.
DR Pfam; PF00261; Tropomyosin; 1.
FT NON_TER 1
SQ SEQUENCE 22 AA; 2611 MW; F86B844608F4175C CRC64;

Query Match 23.7%; Score 27; DB 5; Length 22;
Best Local Similarity 31.2%; Pred. No. 4.3e+03;

Matches 5; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 6 AHKRIPEEKREKLE 21
| | | | |
Db 1 AEQVYKRLQKEYVDRL 16

RESULT 13

09JUT5 PRELIMINARY; PRT; 14 AA.
AC 09JUT5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE B-Raf protein (Fragment).
GN B-RAF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Barnier J.V.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Barnier J.V., Papin C., Eyche A., Lecocq O.;
RT "The mouse B-raf gene encodes multiple protein isoforms with tissue-
specific expression";
RL J. Biochem. 270:23381-23389(1995).
DR EMBL; AJ276308; CAB81556.1; -.
FT NON_TER 1
SQ SEQUENCE 14 AA; 1748 MW; D1E0505C4927F02 CRC64;

Query Match 22.8%; Score 26; DB 11; Length 14;
Best Local Similarity 33.3%; Pred. No. 3.9e+03;
Matches 4; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19
| | | | |
Db 2 EKPEVELQDOR 13

RESULT 14

09R545 PRELIMINARY; PRT; 15 AA.
AC 09R545;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE 30S ribosomal protein (Fragment).
GN Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetia; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_Taxid=1765;
RN [1]
RP SEQUENCE.
RX MEDLINE=94009653; PubMed=8405418;
RA Ohara N., Kimura M., Higashi Y., Yamada T.;
RT Isolation and amino acid sequence of the 30S ribosomal protein S19
from Mycobacterium bovis BCG.";
RL FEBS Lett. 331:9-14(1993).
SQ SEQUENCE 15 AA; 1674 MW; 07A36F018AE355A4 CRC64;

Query Match 22.8%; Score 26; DB 2; Length 15;
Best Local Similarity 33.3%; Pred. No. 4.1e+03;
Matches 4; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 KRIPEEKREK 19
| | | | |
Db 3 KNVPANSRRKAK 14

```
RESULT 15
O9UBKO          PRELIMINARY;      PRT;      15 AA.
ID O9UBKO
AC O9UBKO;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)
DE Renal tissue-nonspecific alkaline phosphatase (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RX MEDLINE=93092315; PubMed=1458595;
RA Nishihara Y., Hayashi Y., Adachi T., Koyama I., Stigbrand T.,
RA Hirano K.,
RT "Chemical nature of intestinal-type alkaline phosphatase in human
RT kidney."
RL Clin. Chem. 38:2539-2542(1992).
SQ SEQUENCE 15 AA; 1931 MW; 9A28FDE13F01F716 CRC64;

Query Match 22.8%; Score 26; DB 4; Length 15;
Best Local Similarity 66.7%; Pred. No. 4.1e+03;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 10 IPEEK 15
   :||:|
Db 2 VPEEK 7
```

Search completed: July 16, 2003, 08:14:20
Job time : 28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 07:56:58 / Search time 35 Seconds
(without alignments)
83.758 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFFAHKKIPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 303745

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A.Geneseq_101002.*
2: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
3: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
4: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
5: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	100.0	22	21	AAV32244
2	44	38.6	20	22	AA85156
3	43	37.7	20	17	AA88078
4	43	37.7	20	22	AA859147
5	42	36.8	19	19	AA48683
6	39	34.2	20	23	AA85587
7	38	33.3	22	22	AB820986
8	38	33.3	22	22	AA68751
9	38	33.3	22	22	AA68751
10	37	32.5	20	23	AA85588

11	36	31.6	13	22	AA899530
12	36	31.6	13	23	AA874746
13	36	31.6	14	22	AA859149
14	36	31.6	16	22	AA899524
15	36	31.6	16	22	AA859148
16	36	31.6	17	22	AA859533
17	36	31.6	18	21	AAV52610
18	36	31.6	19	12	AAW14811
19	36	31.6	19	17	AA89965
20	36	31.6	19	22	AA899534
21	36	31.6	21	16	AA873887
22	36	31.6	21	16	AA864349
23	36	31.6	21	21	AA852640
24	36	31.6	21	21	AA814624
25	36	31.6	21	22	AA854887
26	36	31.6	22	9	AA82835
27	36	31.6	22	16	AA864350
28	36	31.6	22	21	AA852641
29	36	31.6	22	21	AA854888
30	36	31.6	22	22	AA854888
31	35	30.7	10	20	AAV06448
32	35	30.7	20	17	AA888076
33	34	29.8	14	22	AA892596
34	34	29.8	18	17	AAW06514
35	34	29.8	19	18	AAW00941
36	34	29.8	20	16	AA82651
37	34	29.8	20	21	AA820799
38	34	29.8	21	16	AA892636
39	34	29.8	21	18	AAW14600
40	33.5	29.4	15	22	AA859157
41	33	28.9	14	19	AAW69289
42	33	28.9	15	16	AA82650
43	33	28.9	19	20	AAW67988
44	33	28.9	20	19	AAW67988
45	33	28.9	20	21	AA814755

ALIGNMENTS

RESULT 1	
AAV32244	
ID AAV32244 standard; Peptide; 22 AA.	
XX	
AC AAV32244;	
XX	
DT 15-FEB-2000 (first entry)	
XX	
DE Human integrin subunit alpha-10 cytoplasmic domain peptide.	
XX	
KW Integrin alpha-10; ISA10; human; trauma; rheumatoid arthritis;	
KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;	
KW inflammation; therapy; cartilage; chondrocyte; osteoblast;	
KW fibroblast; vaccine; marker.	
XX	
OS Homo sapiens.	
XX	
PN W09951639-A1.	
XX	
PD 14-OCT-1999.	
XX	
PF 31-MAR-1999; 99WO-SE00544.	
XX	
PR 02-APR-1998; 98SE-0001164.	
PR 28-JAN-1999; 99SE-0000319.	
XX	
PA (ACTI-) ACTIVE BIOTECH AB.	
XX	
PI Lundgren-Akerlund E.	
XX	
DR WPI; 2000-052639/04.	
XX	
PT New isolated integrin subunit alpha-10, used as a marker or target	

Human Hsp70B' anti
Transcription fact
Alpha IIB beta 3
Human Hsp70B' anti
Alpha IIB beta 3
Hsp70B' immunogeni
v-myb encoded onco
myb oncogene prote
Clan12 B-cell epit
Hsp70B' immunogeni
HIV-1 envelope pol
DP-178 homologue 1
T20/DP178 peptide
HIV-1 isolate LAI
Anti-HIV peptide D
Whooping cough ant
DP-178 homologue 2
T20/DP178 peptide
HIV-1 isolate LAI
Anti-HIV peptide D
CI esterase and Fa
Cytoplasmic integr
Human peptide #871
HIV-1 matrix prote
Acidix acidix ext
HML-1 alpha-E chat
PEST sequence pep
HML-1 alpha-E chat
Synthetic ubiqit1
Alpha IIB beta 3
Maspin epitope. H
Fragment of human
Human Fas peptide
Human Fas epitope,

PT molecule for cells during development, regeneration and pathological
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT inflammation

PS Claim 21; Page 53; 90pp; English.

CC This sequence represents a fragment of novel human chondrocyte
CC integrin subunit alpha-10 (Isa10, see AAY32242), corresponding to
CC the C-terminal cytoplasmic domain of the protein. The invention
CC relates to a recombinant or isolated integrin heterodimer
CC comprising the alpha10 subunit in association with subunit beta
CC (especially beta-1). The integrin heterodimer, or the subunit
CC alpha-10, or a fragment of it such as the present sequence, can be
CC used as a marker or target of all types of cells, e.g. of
CC chondrocytes, osteoblasts and fibroblasts. They can also be used:
CC for treating pathological conditions involving Isa10, such as
CC damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
CC for detecting the formation of cartilage during embryonal
CC development, physiological or therapeutic repair of cartilage,
CC or detecting regeneration of cartilage or chondrocytes during
CC transplantation of cartilage or chondrocytes; for selection and
CC analysis or for sorting, isolating or purification of chondrocytes
CC and for in vitro studies of differentiation of chondrocytes; and as
CC a target for anti-adhesive drugs or molecules in tendon, ligament,
CC skeletal muscle or other tissues where adhesion impairs the function
CC of the tissue (all claimed). Isa10 binding entities can be used to
CC determine the differentiation-state of cells during embryonic
CC development, angiogenesis or development of cancer, in pathological
CC conditions such as rheumatoid arthritis, osteoarthritis or cancer,
CC in tissue regeneration or in therapeutic and physiological repair
CC of cartilage (claimed). A vaccine comprising the integrin
CC heterodimer or subunit alpha-10 is also claimed. Isa10
CC polynucleotides, vectors, host cells and methods of producing
CC recombinant Isa10 are also claimed.

SQ Sequence 22 AA;

Query Match 100.0%; Score 114; DB 21; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.7e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KLGFAHKKIPPEEKREKLEQ 22
Db 1 KLGFAHKKIPPEEKREKLEQ 22

RESULT 2
AAB59156
ID AAB59156 standard; peptide; 20 AA.

AC AAB59156;

DT 21-MAR-2001 (first entry)

DE Alpha IIB beta 3 binding motif peptide #10.

KW Hepatid repeat; transmembrane domain; cytoplasmic; integrin;
KW inflammation; thrombosis; malignancy.

OS Synthetic.

PN WO200073341-A1.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14656.

PR 27-MAY-1999; 99US-0320907.

PA (SCRI) SCRIPPS RES INST.

PI Ginsberg MH, Pfaff M;

DR WPI; 2001-041143/05.

XX Polypeptides useful in construction of structural models for.
PT Identifying therapeutic compounds, comprises series of heptad repeats
PT that mimic a transmembrane domain and cytoplasmic domain attached to
PT the repeats

PS Disclosure; Page 13; 36pp; English.

CC The present invention relates to a peptide with a series of
CC heptad-repeats that mimic a transmembrane domain and a selected
CC cytoplasmic domain attached to the heptad repeats. The invention
CC is useful for evaluating structure and activity of a selected
CC occupied and clustered transmembrane protein with the selected
CC cytoplasmic domain and for identifying therapeutic compounds. It
CC is also useful for identifying a cytoplasmic domain binding partner.
CC It is may be used to study protein interactions with transmembrane
CC proteins such as integrin, which can be used to treat conditions in
CC which over activity of integrins is involved, such as inflammation,
CC thrombosis and malignancy.

SQ Sequence 20 AA;

Query Match 38.6%; Score 44; DB 22; Length 20;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 KLGFAHKKIPPEEKRE 18
Db 1 KVGFFKRNPPLEBAEE 18

RESULT 3

ID AAR88078 standard; peptide; 20 AA.

AC AAR88078;

DT 30-JUL-1996 (first entry)

DE Integrin alpha-IIB cytoplasmic region.

KW Helical; soluble; amphiphilic; integrin; chimera; inflammation;
KW inhibition; thrombosis; malignancy; transmembrane receptor.

OS Homo sapiens.

PN WO9534641-A1.

PD 21-DEC-1995.

PE 13-JUN-1995; 95WO-US07542.

PR 15-JUN-1994; 94US-0260514.

PA (SCRI) SCRIPPS RES INST.

PI Chen Y, Ginsberg MH, Kent SBH, Muir TW, O'Toole TE;

PI Williams MJ;

DR WPI; 1996-049670/05.

XX New soluble protein models of transmembrane receptors - have helical
PT amphiphilic sequence to represent transmembrane domain, also
PT chimeric integrin proteins for inhibiting human cellular integrins
XX Disclosure; Fig 9; 10pp; English.

CC AAR88078 represents the human integrin alpha-IIB cytoplasmic domain.
CC It is used to produce a Tac/integrin chimera having the cytoplasmic
CC domain as above covalently linked to the extracellular and transmembrane
CC domain of the Tac subunit of the human interleukin-2 receptor. Chimeric
CC integrins can be used to block the activity of natural integrins in

PF 30-JAN-2001; 2001WO-US00668.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI, 2001-488900/53.
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 CC
 PS Example 4; SEQ ID NO: 29057; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 SQ Sequence 22 AA;

Query Match 33.3%; Score 38; DB 22; Length 22;
 Best Local Similarity 61.5%; Pred. No. 1.2e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 7 HKKIPEEKREK 19
 :|||:|
 Db 3 NKKKEEEEEEEK 15

RESULT 9
 AAM16576
 ID AAM16576 standard; Protein; 22 AA.
 XX
 AC AAM16576;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Peptide #3010 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI, 2001-488901/53.
 XX
 DR

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 XX Claim 27; SEQ ID No 21402; 487bp; English.
 PS
 CC The present invention relates to human single exon nucleic acid probes
 CC (SNP: see A110068-A128459). The present sequence is a peptide encoded
 CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 22 AA;

Query Match 33.3%; Score 38; DB 22; Length 22;
 Best Local Similarity 61.5%; Pred. No. 1.2e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 Oy 7 HKKIPEEKREK 19
 :|||:|
 Db 3 NKKKEEEEEEEK 15

RESULT 10
 AAU85588
 ID AAU85588 standard; Peptide; 20 AA.
 XX
 AC AAU85588;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Lung tumour protein L548S peptide #23.
 XX
 DE Lung tumour; cancer; T cell; immune response stimulator;
 XX
 KW cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN WO200204514-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 10-JUL-2001; 2001WO-US22058.
 XX
 PR 11-JUL-2000; 2000US-0614124.
 PR 29-AUG-2000; 2000US-0651563.
 PR 08-SEP-2000; 2000US-0658824.
 PR 26-SEP-2000; 2000US-0671325.
 PR 06-OCT-2000; 2000US-0677419.
 PR 30-OCT-2000; 2000US-0702705.
 PR 13-DEC-2000; 2000US-0736457.
 PR 03-MAY-2001; 2001US-0849626.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
 PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;
 PI McNabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;
 XX WPI, 2002-164634/21.
 DR
 XX Novel polynucleotide encoding a lung tumour polypeptide useful for
 PT stimulating and/or expanding T cells specific for a tumour protein
 XX
 PS Claim 2; SEQ ID No 1856; 223bp; English.
 XX
 CC The invention describes an isolated polynucleotide and polypeptide

CC produced are useful for inhibiting the progression of neoplastic
 CC diseases. The invention relates to the field of gene therapy and is
 CC directed toward methods for producing peptide-lipid-polynucleotide
 CC complexes suitable for delivery of polynucleotides. The encapsulated
 CC molecules display therapeutic efficacy in eradicating solid tumours
 CC including but not limited to breast carcinoma or prostate carcinoma.
 CC ABB74235 co ABB74255 are used in the exemplification of the present
 CC invention.

CC Sequence 13 AA;

Query Match 31.6%; Score 36; DB 23; Length 13;
 Best Local Similarity 63.6%; Pred. No. 1.4e+02;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 PEEKREKLE 21
 ||:||||:
 Db 3 PEKEKRIKLE 13

RESULT 13

AAB59149
 ID AAB59149 standard; peptide; 14 AA.

AC AAB59149;

DT 21-MAR-2001 (first entry)

DE Alpha IIB beta 3 binding motif peptide #3.

KM Heptad repeat; transmembrane domain; cytoplasmic; integrin;
 inflammation; thrombosis; malignancy.

OS Synthetic.

PN WO200073341-A1.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14656.

PR 27-MAY-1999; 99US-0320907.

PA (SCRI) SCRIPPS RES INST.

PI Gineberg MH, Pfaff M;

DR WPI; 2001-041143/05.

PT Polypeptides useful in construction of structural models for
 PT identifying therapeutic compounds, comprises series of heptad repeats
 PT that mimic a transmembrane domain and cytoplasmic domain attached to
 PT the repeats

PS Disclosure; Page 13; 36pp; English.

CC The present invention relates to a peptide with a series of
 CC heptad-repeats that mimic a transmembrane domain and a selected
 CC cytoplasmic domain attached to the heptad repeats. The invention
 CC is useful for evaluating structure and activity of a selected
 CC occupied and clustered transmembrane protein with the selected
 CC cytoplasmic domain and for identifying therapeutic compounds. It
 CC is also useful for identifying a cytoplasmic domain binding partner.
 CC It is may be used to study protein interactions with transmembrane
 CC proteins such as integrin, which can be used to treat conditions in
 CC which over activity of integrins is involved, such as inflammation,
 CC thrombosis and malignancy.

CC Sequence 14 AA;

Query Match 31.6%; Score 36; DB 22; Length 14;
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KLGFPFAKKIPEEE 14
 ||:||||:
 Db 1 KVGFFKRRRPPLEE 14

RESULT 14

AAB9524
 ID AAB9524 standard; peptide; 16 AA.

AC AAB9524;

DT 04-SEP-2001 (first entry)

DE Human Hsp70B' antigenic peptide #2.

KM Human; Hsp70B'; detection; stress-inducible protein; antigenic;
 KM epitope; immunogenic; hypertension; oncology; organ transplantation;
 KM ischaemia; trauma; infection; inflammation; fever; heart disease;
 KM autoimmune disorder; neoplasia.

OS Homo sapiens.

PN WO200142423-A2.

PD 14-JUN-2001.

PF 07-DEC-2000; 2000WO-US33341.

PR 07-DEC-1999; 99US-0169535.

PA (STRE-) STRESSGEN BIOTECHNOLOGIES CORP.

PI Boux HA, Wong GS, Rodriguez H;

DR WPI; 2001-381658/40.

PT New antibody that specifically binds to heat shock protein 70B' or to
 PT HSP70B' peptides, useful for determining whether cell has been exposed
 PT to stressful environment or stressful substance
 PS Claim 1; Page 56; 64pp; English.

CC The present invention describes an antibody (I) that specifically binds
 CC to the human heat shock protein (Hsp70B', or to Hsp70B' peptides. (I)
 CC is useful for determining whether a cell has been exposed to a stressful
 CC environment or a stressful substance by an immunoassay in which proteins
 CC in the cell or proteins extracted from the cell are exposed to (I). (I)
 CC can be used to distinguish the expression of Hsc70/Hsp70 proteins, which
 CC occur while a cell is functioning normally and when it is responding to
 CC stress, and the Hsp70B' protein, which is only produced in response to
 CC stress. A positive reaction to Hsp70B' antibodies provides an indication
 CC of the general state of the health of the organism, in which that cell
 CC resides. The anti-Hsp70B' antibodies can be used to monitor the protein
 CC levels of a specific, sensitive, native biomarker (Hsp70B') in in vitro
 CC bioassays using human cell lines to evaluate the toxicity of chemical
 CC compounds. Monitoring Hsp70B' is useful for evaluating the fitness of
 CC healthy individuals, in the care of patients who are in remission from
 CC an autoimmune, chronic condition or neoplasia, to evaluate amniotic
 CC fluids or samples of the placenta and to assess new born infants who are
 CC at risk. (I) is also useful for evaluation, diagnosis, prognosis and
 CC continued monitoring of specific disease conditions such as hypertension,
 CC oncology, organ transplantation, ischaemia, trauma, infection,
 CC inflammation, fever, heart disease and autoimmune disorders. The present
 CC sequence represents a specifically claimed Hsp70B' peptide from the
 CC present invention.

CC Sequence 16 AA;

Query Match 31.6%; Score 36; DB 22; Length 16;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 9 KIPEEKRE 17
 |||||:
 Db 3 KIPEEDRRK 11

RESULT 15

AAB59148
 ID AAB59148 standard; peptide; 16 AA.

XX AAB59148;
 AC AAB59148;

DT 21-MAR-2001 (first entry)

XX Alpha IIB beta 3 binding motif peptide #2.

DE Heprad repeat; transmembrane domain; cytoplasmic; integrin;
 XX inflammation; thrombosis; malignancy.

OS Synthetic.

XX WO200073341-A1.

PD 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US14656.

XX 27-MAY-1999; 99US-0320907.

XX (SCRI) SCRIPPS RES INST.

XX Ginsberg MH, Pfaff M;

DR MPI; 2001-041143/05.

XX
 PT Polypeptides useful in construction of structural models for
 PT identifying therapeutic compounds, comprises series of heptad repeats
 PT that mimic a transmembrane domain and cytoplasmic domain attached to
 PT the repeats -

PS Disclosure; Page 13; 36pp; English.

XX
 CC The present invention relates to a peptide with a series of
 CC heptad-repeats that mimic a transmembrane domain and a selected
 CC cytoplasmic domain attached to the heptad repeats. The invention
 CC is useful for evaluating structure and activity of a selected
 CC occupied and clustered transmembrane protein with the selected
 CC cytoplasmic domain and for identifying therapeutic compounds. It
 CC is also useful for identifying a cytoplasmic domain binding partner.
 CC It is used to study protein interactions with transmembrane
 CC proteins such as integrin, which can be used to treat conditions in
 CC which over activity of integrins is involved, such as inflammation,
 CC thrombosis and malignancy.

XX Sequence 16 AA;

Query Match 31.6%; Score 36; DB 22; Length 16;
 Best Local Similarity 50.0%; Pred. No. 1.7e+02;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 KLGFPAHKKIPPEE 14
 .:|:|:|:|:|:|:
 Db 1 KVGFFKRNRPPLPE 14

Search completed: July 16, 2003, 08:13:28
 Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:14:28 ; Search time 21 Seconds
(without alignments)
124.415 Million cell updates/sec

Title: US-09-647-544-7
Perfect score: 114
Sequence: 1 KLGFFAHKKIPEEKKEKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 segs, 118759770 residues

Total number of hits satisfying chosen parameters: 96116

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCIT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCITUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	38.6	19	US-10-101-487-67	Sequence 67, Appl
2	44	38.6	20	US-09-320-907B-24	Sequence 24, Appl
3	43	37.7	20	US-09-320-907B-15	Sequence 15, Appl
4	39	34.2	20	US-09-736-457-1855	Sequence 1855, Ap
5	39	34.2	20	US-09-902-941-1855	Sequence 1855, Ap
6	39	34.2	20	US-09-849-626-1855	Sequence 1855, Ap
7	39	34.2	20	US-10-017-754-1855	Sequence 1855, Ap
8	38	33.3	22	US-09-864-761-36284	Sequence 36284, A
9	37	33.5	20	US-09-736-457-1856	Sequence 1856, Ap
10	37	33.5	20	US-09-902-941-1856	Sequence 1856, Ap
11	37	32.5	20	US-09-849-626-1856	Sequence 1856, Ap
12	37	32.5	20	US-10-017-754-1856	Sequence 1856, Ap
13	36	31.6	13	US-09-733-179A-5	Sequence 5, Appl
14	36	31.6	13	US-09-876-904A-510	Sequence 510, App
15	36	31.6	14	US-09-320-907B-17	Sequence 17, Appl
16	36	31.6	14	US-09-733-179A-15	Sequence 15, Appl
17	36	31.6	16	US-09-320-907B-16	Sequence 16, Appl
18	36	31.6	16	US-09-733-179A-4	Sequence 4, Appl
19	34	29.8	9	US-10-059-720-4	Sequence 4, Appl

20	34	29.8	20	US-09-880-132-1	Sequence 1, Appl
21	34	29.8	20	US-09-880-149-1	Sequence 1, Appl
22	34	29.8	20	US-10-038-045-15	Sequence 15, Appl
23	34	29.8	21	US-10-057-789-231	Sequence 231, App
24	34	29.8	21	US-10-212-628-231	Sequence 231, App
25	33.5	29.4	15	US-09-320-907B-25	Sequence 25, Appl
26	33.5	29.4	20	US-10-142-935-6	Sequence 6, Appl
27	33	28.9	19	US-09-984-245-306	Sequence 306, App
28	33	28.9	19	US-09-966-262-306	Sequence 306, App
29	33	28.9	19	US-09-983-966-306	Sequence 306, App
30	33	28.9	19	US-10-143-090-306	Sequence 306, App
31	33	28.9	20	US-10-216-484-30	Sequence 30, Appl
32	33	28.9	21	US-09-805-301-81	Sequence 81, Appl
33	33	28.1	18	US-09-865-989-249	Sequence 249, App
34	32	28.1	18	US-10-099-574A-249	Sequence 249, App
35	31	27.2	12	US-09-876-904A-300	Sequence 300, App
36	31	27.2	16	US-09-832-355A-40	Sequence 40, Appl
37	31	27.2	16	US-10-001-073-8	Sequence 8, Appl
38	31	27.2	16	US-10-225-567A-948	Sequence 948, App
39	31	27.2	18	US-10-116-391-32	Sequence 32, Appl
40	31	27.2	20	US-09-945-917-32	Sequence 32, Appl
41	31	27.2	20	US-09-777-558-1	Sequence 1, Appl
42	31	27.2	21	US-09-974-879-188	Sequence 388, App
43	31	27.2	21	US-09-305-736-388	Sequence 388, App
44	30	26.3	13	US-09-919-124-26	Sequence 26, Appl
45	30	26.3	13	US-09-919-124-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-10-101-487-67
Sequence 67, Application US/10101487
Patent No. US20020169125A1
GENERAL INFORMATION:
APPLICANT: LEUNG, DAVID W.
APPLICANT: BERGMAN, PHILIP A.
APPLICANT: LOFOUST, ALAN
APPLICANT: PIETZ, GREGORY E.
APPLICANT: TOMPKINS, CHRISTOPHER K.
APPLICANT: WAGGNER JR., DAVID W.
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: 077319/0329
CURRENT APPLICATION NUMBER: US/10/101,487
CURRENT FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: 60/277,705
PRIOR FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 116
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 67
LENGTH: 19
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
US-10-101-487-67

Query Match 38.6%; Score 44; DB 9; Length 19;
Best Local Similarity 57.1%; Pred. No. 9.5;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 KIPPEEKKEKLEQ 22
DB 2 KMPEEEEEEEEE 15

RESULT 2
US-09-320-907B-24
Sequence 24, Application US/09320907B
Patent No. US20020198360A1
GENERAL INFORMATION:

```
;; GENERAL INFORMATION:
;; APPLICANT: Bangur, Chaitanya
;; APPLICANT: Fanger, Gary
;; APPLICANT: Wang, Aijun
;; APPLICANT: Wang, Tongtong
;; APPLICANT: Switzer, Anne
```

```

; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1855
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1855

Query Match          34.2%; Score 39; DB 9; Length 20;
Best Local Similarity 53.8%; Pred. No. 48;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 6 AHKKIPEEEKKEE 18
    |||:||||:
Db 8 ARKKVEEDEDEEE 20

RESULT 7
US-10-017-754-1855
; Sequence 1855, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Derrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedavick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1855
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1855

Query Match          34.2%; Score 39; DB 9; Length 20;
Best Local Similarity 53.8%; Pred. No. 48;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 6 AHKKIPEEEKKEE 18
    |||:||||:
Db 8 ARKKVEEDEDEEE 20

RESULT 8
US-09-864-761-36284
; Sequence 36284, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; GENE EXPRESSION ANALYSIS BY MICROARRAY
```

```

; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36284
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007158.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
US-09-864-761-36284

Query Match          33.3%; Score 38; DB 10; Length 22;
Best Local Similarity 61.5%; Pred. No. 72;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 7 HKKIPPEEKKEE 19
    |||:||||:
Db 3 NKKEEKEEKEE 15

RESULT 9
US-09-736-457-1856
; Sequence 1856, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedavick, Tom
; APPLICANT: Carter, Derrick
; APPLICANT: Retter, Marc
```

```

; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-1856

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 10
US-09-902-941-1856
; Sequence 1856, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-1856
```

```

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 11
US-09-849-626-1856
; Sequence 1856, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
```

```

; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-1856

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 12
US-10-017-754-1856
; Sequence 1856, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1856
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-754-1856
```

```

Query Match      32.5%; Score 37; DB 9; Length 20;
Best Local Similarity 40.0%; Pred. No. 89;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY      8 KKIPEEKREKLEQ 22
DB      2 KKVEEDEDEQEEREE 16
```

```

RESULT 13
US-09-733-179A-5
; Sequence 5, Application US/09733179A
; Publication No. US20030073160A1
; GENERAL INFORMATION:
; APPLICANT: Boux, Heather A.
; APPLICANT: Wong, Geraldine S.
; APPLICANT: Rodriguez, Henry
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING STRESS-INDUCIBLE PROTEINS
; FILE REFERENCE: 12071-006001
; CURRENT APPLICATION NUMBER: US/09/733,179A
```

```
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: WO US00/33341
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,535
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-733-179A-5

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 13;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 9 KPEEKRE 17
Db 3 KPEEDRRK 11

RESULT 14
US-09-876-904A-510
; Sequence 510, Application US/09876904A
; Publication No. US20030072794A1
; GENERAL INFORMATION:
; APPLICANT: BOULIKAS, TENI
; TITLE OF INVENTION: ENCAPSULATION OF PLASMID DNA (LIPGENES TM) AND THERAPEUTIC
; TITLE OF INVENTION: AGENTS WITH NUCLEAR LOCALIZATION SIGNAL/FUSOGENIC PEPTIDE
; FILE REFERENCE: TB-2002.00
; CURRENT APPLICATION NUMBER: US/09/876,904A
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,925
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 510
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: Mouse c-Myb
US-09-876-904A-510

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 13;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 PEEKREKLE 21
Db 3 PEKREKLE 13

RESULT 15
US-09-320-907B-17
; Sequence 17, Application US/09320907B
; Publication No. US20020198360A1
; GENERAL INFORMATION:
; APPLICANT: GINSBERG, MARK H.
; APPLICANT: PFAFF, MARTIN
; TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC DOMAINS OF
; TITLE OF INVENTION: TRANSMEMBRANE RECEPTORS
; FILE REFERENCE: SRI-0006
; CURRENT APPLICATION NUMBER: US/09/320,907B
; CURRENT FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 09/187,236
; PRIOR FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 14
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-320-907B-17

Query Match
Best Local Similarity 31.6%; Score 36; DB 9; Length 14;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KLGFFAKKPEEE 14
Db 1 KVGFFKRNRPLEE 14

Search completed: July 16, 2003, 08:21:30
Job time : 22 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:26:05 ; Search time 14 Seconds
(without alignments)
3457.348 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167
Sequence: 1 MELPVTMLFLPLVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1167	100.0	1167	1 ITAG_HUMAN	O75578 homo sapien
2	13	1.1	285	1 ITAI_CHICK	O90615 gallus gall
3	13	1.1	1151	1 ITAI_HUMAN	P56199 homo sapien
4	13	1.1	1180	1 ITAI_RAT	P18614 rattus norv
5	13	1.1	1189	1 ITAH_HUMAN	O9ukx5 homo sapien
6	11	0.9	687	1 CSTA_HELPJ	O92k47 heliocobact
7	11	0.9	687	1 CSTA_HELPY	P56190 heliocobact
8	11	0.9	1170	1 ITA2_BOVIN	P53710 bos taurus
9	11	0.9	1178	1 ITA2_MOUSE	O62469 mus musculu
10	11	0.9	1181	1 ITA2_HUMAN	P17301 homo sapien
11	10	0.9	1179	1 ITAE_HUMAN	P38570 homo sapien
12	9	0.8	1163	1 ITAL_MOUSE	P24063 mus musculu
13	9	0.8	1170	1 ITAL_HUMAN	P20701 homo sapien
14	8	0.7	76	1 CD24_MOUSE	P24077 mus musculu
15	8	0.7	76	1 CD24_RAT	O07490 rattus norv
16	8	0.7	80	1 CD24_HUMAN	P56503 homo sapien
17	8	0.7	108	1 PRVB_MERBI	P56503 merluccius
18	8	0.7	179	1 IL22_HUMAN	O9gpx6 homo sapien
19	8	0.7	234	1 EMB2_CAVPO	P35709 cavia porce
20	8	0.7	254	1 PSAT_YEAST	P40325 saccharomyc
21	8	0.7	256	1 HISU_CAMJE	O46125 campylobact
22	8	0.7	259	1 PSAT_SCHPO	O10329 schizosacch
23	8	0.7	264	1 COMT_RAT	P22734 rattus norv
24	8	0.7	309	1 ALSK_ECOLI	P32718 escherichia
25	8	0.7	423	1 NSMA_HUMAN	O60806 homo sapien
26	8	0.7	470	1 CPBK_MOUSE	O62397 mus musculu
27	8	0.7	491	1 CPBI_RAT	P00167 rattus norv
28	8	0.7	491	1 CPB2_RAT	P04167 rattus norv
29	8	0.7	500	1 CPBA_MOUSE	P12791 mus musculu
30	8	0.7	604	1 ITAB_PAPCY	P53711 papio cynoc
31	8	0.7	643	1 ZN74_HUMAN	O16587 homo sapien
32	8	0.7	942	1 TMK1_ARATH	P43298 arabidopsis
33	8	0.7	959	1 VP2_BTIV3V	O06998 bluetongue

34	8	0.7	1033	1 ITAB_MOUSE	O9qum0 mus musculu
35	8	0.7	1039	1 ITAB_HUMAN	P08514 homo sapien
36	8	0.7	1050	1 ITA5_XENLA	O06274 xenopus lae
37	8	0.7	1146	1 ITAI_DROME	O24247 drosophila
38	8	0.7	1152	1 ITAM_HUMAN	P11215 homo sapien
39	8	0.7	1163	1 ITAX_HUMAN	P20702 homo sapien
40	8	0.7	3176	1 CA36_HUMAN	P12111 homo sapien
41	8	0.6	72	1 DMS4_AGAN	O93222 agalychnis
42	7	0.6	73	1 DMS2_AGAN	O93222 agalychnis
43	7	0.6	74	1 DMS3_PHYBI	P81485 phyllomedus
44	7	0.6	77	1 ACP_FOMTE	P80918 comamonas t
45	7	0.6	79	1 ACP1_RALSO	O8y0j1 ralstonia s
46	7	0.6	79	1 ACE_XYLA	O9pf15 xyella tas
47	7	0.6	97	1 INS_CHICK	P44592 haemophilus
48	7	0.6	107	1 INS_CHICK	P01332 gallus gall
49	7	0.6	110	1 HDEA_ECOLI	P26604 escherichia
50	7	0.6	139	1 TGF_MYXGL	P22618 myxine glut
51	7	0.6	142	1 CDU_YEAST	O06549 saccharomyc
52	7	0.6	142	1 MK_CHICK	P24052 gallus gall
53	7	0.6	144	1 GLE_APLJU	P14393 aplysia jul
54	7	0.6	149	1 RHIC_RHIV	O03315 rhizobium
55	7	0.6	160	1 BIK_HUMAN	O13323 homo sapien
56	7	0.6	170	1 YCR5_MYXXA	O06929 myxococcus
57	7	0.6	171	1 ATRF_SYNPE	P08447 synchococc
58	7	0.6	171	1 YER7_METU	O58982 methanococc
59	7	0.6	178	1 PORC_METU	O57717 methanococc
60	7	0.6	181	1 HYEF_ECOLI	P77423 escherichia
61	7	0.6	198	1 IUP1_ECOLI	P18957 escherichia
62	7	0.6	205	1 RAC1_DICDI	O9gpr2 dictyostei
63	7	0.6	206	1 GPEB_MOUSE	P56400 mus musculu
64	7	0.6	208	1 RL13_SCHPO	O74175 schizosacch
65	7	0.6	212	1 RESD_MYCTU	O50712 mycobacteri
66	7	0.6	214	1 RL40_CAEEL	O09533 caenorhabdi
67	7	0.6	220	1 FGP3_CHICK	P48801 gallus gall
68	7	0.6	221	1 SPFL_HUMAN	O9hcn8 homo sapien
69	7	0.6	225	1 EMBP_MOUSE	O61878 mus musculu
70	7	0.6	225	1 BRX1_HUMAN	O9hbu1 homo sapien
71	7	0.6	227	1 EMBP_RAT	O63189 rattus norv
72	7	0.6	231	1 NH40_CAEEL	O22127 caenorhabdi
73	7	0.6	231	1 Y667_RICPR	O9cxc9 rickettsia
74	7	0.6	233	1 EMB1_CAVPO	P22032 cavia porce
75	7	0.6	235	1 COMB_CLOAB	P27082 clostridium
76	7	0.6	235	1 YOGA_ECOLI	O46831 escherichia
77	7	0.6	240	1 IBP6_HUMAN	P24592 homo sapien
78	7	0.6	243	1 YCGE_HUMAN	P75989 escherichia
79	7	0.6	251	1 TPIS_STRPY	P82478 streptococc
80	7	0.6	251	1 YABF_RHISN	P54429 rhizobium s
81	7	0.6	254	1 BRX1_MOUSE	O9er42 mus musculu
82	7	0.6	256	1 KUKF_HUMAN	O9a125 homo sapien
83	7	0.6	258	1 KPM1_ECOLI	P23889 escherichia
84	7	0.6	258	1 KPM2_ECOLI	P24584 escherichia
85	7	0.6	260	1 NTK_XENLA	P25435 xenopus lae
86	7	0.6	262	1 SP1B_HUMAN	O01892 homo sapien
87	7	0.6	266	1 SYGB_MORCA	P77801 moraxella c
88	7	0.6	269	1 TRBU_AGRU	P54919 vaccinia vi
89	7	0.6	269	1 VI03_VACCC	P20499 vaccinia vi
90	7	0.6	269	1 VI03_VACCV	P12923 vaccinia vi
91	7	0.6	269	1 VI03_VARY	P33000 variola vir
92	7	0.6	270	1 REB8_HUMAN	O00124 homo sapien
93	7	0.6	289	1 ROCI_ARATH	O9anu4 arabidopsis
94	7	0.6	295	1 PNK_PSEAE	O93nu2 pseudomonas
95	7	0.6	307	1 PN2_SYNY3	P73955 synechocyst
96	7	0.6	312	1 CAH4_BOVIN	O95323 bos taurus
97	7	0.6	315	1 YG4W_YEAST	P50085 saccharomyc
98	7	0.6	319	1 Y305_BROME	O8d472 bruceella me
99	7	0.6	321	1 MRAY_ENTRH	O07668 enterococcu
100	7	0.6	324	1 AAIR_CHICK	P49892 gallus gall

ALIGNMENTS

RESULT 1

```

ITAG HUMAN
ID ITAG HUMAN STANDARD; PRT; 1167 AA.
AC 075578; Q9UH28;
AD 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-10 precursor.
GN ITGA10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Articular chondrocytes;
RX MEDLINE=98352078; PubMed=9685391;
RA Camper L., Hellman U., Lundgren-Akerlund E.;
RT "Isolation, cloning, and sequence analysis of the integrin subunit
RT alpha10, a beta1-associated collagen binding integrin expressed on
RT chondrocytes."
RL J. Biol. Chem. 273:20383-20389(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart, and Endothelial cells;
RX MEDLINE=20169197; PubMed=10702680;
RA Lehnert K., Ni J., Leung E., Gough S.M., Morris C.M., Liu D.,
RA Wang S.-X., Langley R., Kriessan G.W.;
RT "The integrin alpha10 subunit: expression pattern, partial gene
RT structure, and chromosomal localization."
RL Cytogenet. Cell Genet. 87:238-244(1999).
CC -1- FUNCTION: INTEGRIN ALPHA-10/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-10
CC ASSOCIATES WITH BETA-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSIONS IN
CC MUSCLE AND HEART. FOUND IN ARTICULAR CARTILAGE.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFPA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF074015; AAC31952.1; -
CC EMBL; AF112345; AAF21944.1; -
CC EMBL; AF172723; AAF61638.1; -
CC HSSP; P17301; 1A0X.
CC Genew; HGNC.6135; ITGA10.
CC MIM; 604042; -
CC InterPro; IPR000413; Integrin_alpha.
CC InterPro; IPR002035; WVF_A.
CC Pfam; PF00092; wva; 1.
CC Pfam; PF01839; FG-GAP; 5.
CC PRINTS; PR01185; INTEGRINA.
CC PRINTS; PR00453; WVFADOMAIN.
CC SMART; SM00191; Int_alpha; 5.
CC SMART; SM00327; WVA; 1.
CC PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
CC PROSITE; PS00234; WVF_A; 1.
CC Interim; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
CC Signal; Repeat; Calcium; Magnesium.
CC SIGNAL; Repeat; 1 22
CC CHAIN 23 1167
CC DOMAIN 23 1122
CC TRANSMEM 1123 1145
CC DOMAIN 1146 1167

```

```

FT REPEAT 38 97
FT REPEAT ? ?
FT DOMAIN 167 350
FT REPEAT 365 427
FT REPEAT 428 482
FT REPEAT 483 545
FT REPEAT 546 605
FT REPEAT 608 660
FT DOMAIN 1134 1140
FT CA_BIND 494 502
FT CA_BIND 558 566
FT CA_BIND 620 628
FT DISULFID 76 86
FT DISULFID 666 675
FT DISULFID 681 736
FT DISULFID 789 795
FT CARBOHYD 98 98
FT CARBOHYD 224 234
FT CARBOHYD 336 336
FT CARBOHYD 364 364
FT CARBOHYD 733 733
FT CARBOHYD 763 763
FT CARBOHYD 839 839
FT CARBOHYD 921 921
FT CARBOHYD 1011 1011
FT CARBOHYD 1018 1018
FT CARBOHYD 1039 1039
FT CONFLICT 844 844
FT CONFLICT 909 909
FT CONFLICT 926 926
SQ SEQUENCE 1167 AA; 127573 MW; AFD3ALC25CLAEN0 CRC64;

Query Match 100.0%; Score 1167; DB 1; Length 1167;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MELPVTHLFLPLVLTGLCSPPNIDHHPRLPFPPEPAEFGYSVLQHYGGGQRMALVGA 60
1 MELPVTHLFLPLVLTGLCSPPNIDHHPRLPFPPEPAEFGYSVLQHYGGGQRMALVGA 60
1 MELPVTHLFLPLVLTGLCSPPNIDHHPRLPFPPEPAEFGYSVLQHYGGGQRMALVGA 60
61 PWDGSGRRGDVYRCVPGAHNAPCAKGLGIDYLGNSHPAYMHGMSLLETGDSGG 120
61 PWDGSGRRGDVYRCVPGAHNAPCAKGLGIDYLGNSHPAYMHGMSLLETGDSGG 120
61 PWDGSGRRGDVYRCVPGAHNAPCAKGLGIDYLGNSHPAYMHGMSLLETGDSGG 120
121 FMAACPLMSRACSSVFSSGICARVDASFQDGLAPTAORCPPTYMDVYVLDGNSIYP 180
121 FMAACPLMSRACSSVFSSGICARVDASFQDGLAPTAORCPPTYMDVYVLDGNSIYP 180
121 FMAACPLMSRACSSVFSSGICARVDASFQDGLAPTAORCPPTYMDVYVLDGNSIYP 180
181 MSEVOTFLRLVGLKFLIPEQIQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240
181 MSEVOTFLRLVGLKFLIPEQIQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240
181 MSEVOTFLRLVGLKFLIPEQIQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240
181 MSEVOTFLRLVGLKFLIPEQIQVGLVOYGSPVHNSLGPRTKEEVVRAAKNLSREG 240
241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKACEAGRVRY 300
241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKACEAGRVRY 300
241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKACEAGRVRY 300
241 RETKTAQAIMVACTGFSQSHGPREARLLVVVTGDSHSGEELPALKACEAGRVRY 300
301 GIAVLGHLRRQDSSFLREIRITASDPDRFFFNVTDEAALTDIVDALGRIFGLEGS 360
301 GIAVLGHLRRQDSSFLREIRITASDPDRFFFNVTDEAALTDIVDALGRIFGLEGS 360
301 GIAVLGHLRRQDSSFLREIRITASDPDRFFFNVTDEAALTDIVDALGRIFGLEGS 360
301 GIAVLGHLRRQDSSFLREIRITASDPDRFFFNVTDEAALTDIVDALGRIFGLEGS 360
361 HAENSSFGLEMSQIGFSTHRLKQDILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEEF 420
361 HAENSSFGLEMSQIGFSTHRLKQDILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEEF 420
361 HAENSSFGLEMSQIGFSTHRLKQDILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEEF 420
361 HAENSSFGLEMSQIGFSTHRLKQDILFGMGAYDMGSSVLMLEGGHLLFPPRMALDEEF 420
421 PPALGNHAAVGYSSVSWMLRGRRFLFSGAPRPFHRGKVAIFOLKKGAARVAOSLGE 480
421 PPALGNHAAVGYSSVSWMLRGRRFLFSGAPRPFHRGKVAIFOLKKGAARVAOSLGE 480
421 PPALGNHAAVGYSSVSWMLRGRRFLFSGAPRPFHRGKVAIFOLKKGAARVAOSLGE 480
421 PPALGNHAAVGYSSVSWMLRGRRFLFSGAPRPFHRGKVAIFOLKKGAARVAOSLGE 480
481 QIGSYFGESELPLDTRDGTDTDVLLVAAPMFLGPONKETGRVYVYLVGQOSLLTQGTLO 540
481 QIGSYFGESELPLDTRDGTDTDVLLVAAPMFLGPONKETGRVYVYLVGQOSLLTQGTLO 540
481 QIGSYFGESELPLDTRDGTDTDVLLVAAPMFLGPONKETGRVYVYLVGQOSLLTQGTLO 540
481 QIGSYFGESELPLDTRDGTDTDVLLVAAPMFLGPONKETGRVYVYLVGQOSLLTQGTLO 540
541 PEPDQARFPGAMGALPDLNODGPAVAVGAPLEDHGQALYLHGTQSGVAPHPAQR1A 600

```



```

Db      ||||| 541 PPPPDARGFAMGALPDLNODGFADVAAGAPLEBCHOGALYLHYGTSGVPHPAORLA 600
Qy      ||||| 601 AAMSMHALSYFGSRVDGRDLDDGDDLVDAVAGNOGAILSSRPVYHLPSLEVPOLAS 660
Db      ||||| 601 AAMSMHALSYFGSRVDGRDLDDGDDLVDAVAGNOGAILSSRPVYHLPSLEVPOLAS 660
Qy      ||||| 661 VVORDCRRRGQBAVCLTAALCFQVTSRTPGRWDHGFYKMFSTALEDWTAAGAAADGSGQ 720
Db      ||||| 661 VVORDCRRRGQBAVCLTAALCFQVTSRTPGRWDHGFYKMFSTALEDWTAAGAAADGSGQ 720
Qy      ||||| 721 RLSRRLRLSVGNVTCQHLHFVLDTSVLRPVALTVTVALNTTKPGVNLNEGSPTSIQ 780
Db      ||||| 721 RLSRRLRLSVGNVTCQHLHFVLDTSVLRPVALTVTVALNTTKPGVNLNEGSPTSIQ 780
Qy      ||||| 781 KLVPSKDCGPDNECVTDVLQVNMDIRSKAPFVVGGRKRVLVTLENRKNAVNT 840
Db      ||||| 781 KLVPSKDCGPDNECVTDVLQVNMDIRSKAPFVVGGRKRVLVTLENRKNAVNT 840
Qy      ||||| 841 SLSIFSRNLHLASLTPQRESPIKVECAAPSAHARLCVGHVPFGAKVTFLEEFSC 900
Db      ||||| 841 SLSIFSRNLHLASLTPQRESPIKVECAAPSAHARLCVGHVPFGAKVTFLEEFSC 900
Qy      ||||| 901 SLLSQVRKCLTASDLSLRNGTLQENTQTSAYIQYEPHLLFSSBSTLHRYEHPYGT 960
Db      ||||| 901 SLLSQVRKCLTASDLSLRNGTLQENTQTSAYIQYEPHLLFSSBSTLHRYEHPYGT 960
Qy      ||||| 961 PVGPGEFPTLRVNLGCVVSGLLISALLPAVHAGVYFLSLQVITNNASCIVONT 1020
Db      ||||| 961 PVGPGEFPTLRVNLGCVVSGLLISALLPAVHAGVYFLSLQVITNNASCIVONT 1020
Qy      ||||| 1021 EPPGPVHPEELOHTNRLNGSNTQCVVACHLGOLAKGTEVSGLLRLVHNEFFRARK 1080
Db      ||||| 1021 EPPGPVHPEELOHTNRLNGSNTQCVVACHLGOLAKGTEVSGLLRLVHNEFFRARK 1080
Qy      ||||| 1081 SLLTVSTFELGTGEEGVVLQLTASRWSESLLEVQTRPILISLWILIGVLGGLLALL 1140
Db      ||||| 1081 SLLTVSTFELGTGEEGVVLQLTASRWSESLLEVQTRPILISLWILIGVLGGLLALL 1140
Qy      ||||| 1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167
Db      ||||| 1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167

RESULT 2
ITAL_CHICK
ID      ||||| 15-JUL-1998 (Rel. 36, Created)
AC      ||||| 15-JUL-1998 (Rel. 36, Last sequence update)
DT      ||||| 15-JUN-2002 (Rel. 41, Last annotation update)
DE      ||||| Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a)
CN      ||||| ITGA1.
OS      ||||| Gallus gallus (Chicken).
OC      ||||| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      ||||| Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianine;
OC      ||||| Gallus.
OX      ||||| NCBI_TaxID=9031;
RN      ||||| [1]
RP      ||||| SEQUENCE FROM N.A.
RC      ||||| TISSUE=gizzard;
RX      ||||| MEDLINE=94357930; PubMed=7521332;
RA      ||||| Kern A., Briesewitz R., Bank I., Marcantonio E.E.;
RT      ||||| "The role of the I domain in ligand binding of the human integrin
RT      ||||| alpha 1 beta 1.";
RL      ||||| J. Biol. Chem. 269:22811-22816(1994).
CC      ||||| -I- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC      ||||| COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC      ||||| E-R IN COLLAGEN.
CC      ||||| -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC      ||||| ASSOCIATES WITH BETA-1.
CC      ||||| -I- SUBCELLULAR LOCATION: Type I membrane protein.

```

```

CC      ||||| -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC      ||||| WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      ||||| -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      ||||| -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC      ||||| -----
CC      ||||| This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      ||||| between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      ||||| the European Bioinformatics Institute. There are no restrictions on its
CC      ||||| use by non-profit institutions as long as its content is in no way
CC      ||||| modified and this statement is not removed. Usage by and for commercial
CC      ||||| entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      ||||| or send an email to license@isb-sib.ch).
CC      ||||| -----
CC      ||||| DR EMBL; U10114; AAA59067.1; -.
CC      ||||| DR HSSP; P17301; IAOX.
CC      ||||| DR InterPro; IPR000413; Integrin_alpha.
CC      ||||| DR InterPro; IPR002035; VWF_A.
CC      ||||| DR Pfam; PF00092; vwa; 1.
CC      ||||| DR SMART; SM00327; VWA; 1.
CC      ||||| DR PROSITE; PS00242; INTEGRIN_ALPHA; PARTIAL.
CC      ||||| DR PROSITE; PS50234; VWFA; 1.
CC      ||||| KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane.
CC      ||||| FT NON_TER 1 1
CC      ||||| FT DOMAIN <1 285 EXTRACELLULAR (POTENTIAL).
CC      ||||| FT DOMAIN 66 279 VWFA.
CC      ||||| FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      ||||| FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      ||||| FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      ||||| FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      ||||| FT NON_TER 285 285
CC      ||||| SQ SEQUENCE 285 AA; 31503 MW; 1B05D3246CD5CA7E CRC64;

Query Match 1.1%; Score 13; DB 1; Length 285;
Best Local Similarity 100.0%; Pred. No. 9e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      ||||| 169 VIVLDGNSNIYPM 181
Db      ||||| 65 VIVLDGNSNIYPM 77.

RESULT 3
ITAL_HUMAN
ID      ||||| 01-NOV-1997 (Rel. 35, Created)
AC      ||||| 01-NOV-1997 (Rel. 35, Last sequence update)
DT      ||||| 15-JUN-2002 (Rel. 41, Last annotation update)
DE      ||||| Integrin alpha-1 (Laminin and collagen receptor) (VLA-1) (CD49a).
CN      ||||| ITGA1.
OS      ||||| Homo sapiens (Human).
OC      ||||| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      ||||| Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      ||||| NCBI_TaxID=9606;
RN      ||||| [1]
RP      ||||| SEQUENCE FROM N.A.
RX      ||||| MEDLINE=93155124; PubMed=8428973;
RA      ||||| Briesewitz R., Epstein M.R., Marcantonio E.E.;
RT      ||||| "Expression of native and truncated forms of the human integrin alpha
RT      ||||| 1 subunit.";
RL      ||||| J. Biol. Chem. 268:2989-2996(1993).
CC      ||||| -I- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC      ||||| COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC      ||||| E-R IN COLLAGEN.
CC      ||||| -I- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC      ||||| ASSOCIATES WITH BETA-1.
CC      ||||| -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC      ||||| -I- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC      ||||| WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      ||||| -I- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      ||||| -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC      ||||| -I- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC      ||||| -I- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC      ||||| -I- DATABASE: NAME=PROV; NOTE=CD guide CD49a entry;

```

```

CC      WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd49a.htm".
DR      HSSP; P17301; IAOX.
DR      Genew; HGNC:6134; ITGAL.
DR      MIM; 192968; -.
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam; PF00092; vwa; 1.
DR      Pfam; PF00357; Integrin_A; 1.
DR      Pfam; PF01839; FG-GAP 5.
DR      PRINTS; PR00453; VWFADOMAIN.
DR      SMART; SM00191; Int_alpha; 5.
DR      SMART; SM00327; VWA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; VWF_A; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Repeat; Calcium; Magnesium.
FT      DOMAIN 1 113 EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 114 1136 POTENTIAL.
FT      DOMAIN 1137 1151 CYTOPLASMIC (POTENTIAL).
FT      REPEAT 16 75 FG-GAP 1.
FT      REPEAT 2 2 VWF_A.
FT      DOMAIN 147 360 FG-GAP 3.
FT      REPEAT 349 404 FG-GAP 4.
FT      REPEAT 405 457 FG-GAP 5.
FT      REPEAT 459 520 FG-GAP 6.
FT      REPEAT 540 599 FG-GAP 7.
FT      REPEAT 602 654 POTENTIAL.
FT      CA_BIND 470 478 POTENTIAL.
FT      CA_BIND 552 560 POTENTIAL.
FT      CA_BIND 614 622 POTENTIAL.
FT      SITE 1139 1142 GEFPR MOTIF.
FT      DISULFID 54 64 BY SIMILARITY.
FT      DISULFID 660 669 BY SIMILARITY.
FT      DISULFID 675 728 BY SIMILARITY.
FT      DISULFID 780 786 BY SIMILARITY.
FT      DISULFID 850 858 BY SIMILARITY.
FT      DISULFID 1002 1034 BY SIMILARITY.
FT      DISULFID 1037 1044 BY SIMILARITY.
FT      CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 313 313 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 390 390 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 504 504 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 671 671 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 752 752 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 812 812 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 855 855 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 880 880 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 946 946 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 980 980 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1055 1055 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1074 1074 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD 1085 1085 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE 1151 AA; 127837 MW; 683F3C1AABF52808 CRC64;
Query Match 1.1%; Score 13; DB 1; Length 1151;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      169 VIVLDGNSNITYPW 181
DB      146 VIVLDGNSNITYPW 158

```

```

RESULT 4
ID      ITAL_RAT      STANDARD;      PRT; 1180 AA.
AC      P18614;
DT      01-NOV-1990 (Rel. 16; Created)
DT      01-NOV-1990 (Rel. 16; Last sequence update)
DT      15-JUN-2002 (Rel. 41; Last annotation update)
DE      Integrin alpha-1 precursor (Laminin and collagen receptor) (VLA-1)
DE      (CD49a).
GN      ITGAL.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_Taxid=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90338125; PubMed=2380249;
RA      Ignatius M.J., Large T.H., Houde M., Tawil J.W., Barton A.,
RA      Esch F., Carbonetto S., Reichardt L.F.;
RT      "Molecular cloning of the rat integrin alpha 1-subunit: a receptor
RT      for laminin and collagen."
RT      J. Cell Biol. 111:709-720 (1990).
[2]
RX      X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 151-364.
RP      MEDLINE=99313197; PubMed=10386626;
RA      Nole M., Pepinsky R.B., Vanyaminov S.Y., Koteliansky V.,
RA      Gotwals P.J., Kaipuaas M.;
RT      "Crystal structure of the alphabeta1 integrin I-domain: insights into
RT      integrin I-domain function."
RT      FEBS Lett. 452:379-385 (1999).
CC      -1- FUNCTION: INTEGRIN ALPHA-1/BETA-1 IS A RECEPTOR FOR LAMININ AND
CC      COLLAGEN. IT RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-
CC      E-R IN COLLAGEN.
CC      -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-1
CC      ASSOCIATES WITH BETA-1.
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWF_A DOMAIN. INTEGRINS
CC      WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC      -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC      -1- SIMILARITY: CONTAINS 1 VWF_A DOMAIN.
CC      -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; X52140; CAA36384.1; -.
DR      FIR; A35854; A35854.
DR      HSSP; P17301; IAOX.
DR      InterPro; IPR000413; Integrin_alpha.
DR      InterPro; IPR002035; VWF_A.
DR      Pfam; PF00092; vwa; 1.
DR      Pfam; PF00357; Integrin_A; 1.
DR      Pfam; PF01839; FG-GAP 5.
DR      PRINTS; PR00453; VWFADOMAIN.
DR      SMART; SM00191; Int_alpha; 5.
DR      SMART; SM00327; VWA; 1.
DR      PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR      PROSITE; PS50234; VWF_A; 1.
KW      Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW      Signal; Repeat; Calcium; Magnesium.
FT      SIGNAL 1 28 INTEGRIN_ALPHA-1.
FT      CHAIN 29 1180 EXTRACELLULAR (POTENTIAL).
FT      DOMAIN 1143 1165 POTENTIAL.
FT      TRANSMEM 1166 1180 CYTOPLASMIC (POTENTIAL).
FT      DOMAIN 1180 1203 FG-GAP 1.

```

```

FT REPEAT ? ? FG-GAP 2.
FT DOMAIN 175 388 VWFA.
FT REPEAT 377 432 FG-GAP 3.
FT REPEAT 433 484 FG-GAP 4.
FT REPEAT 485 565 FG-GAP 5.
FT REPEAT 567 626 FG-GAP 6.
FT REPEAT 629 681 FG-GAP 7.
FT CA_BIND 497 505 POTENTIAL.
FT CA_BIND 579 587 POTENTIAL.
FT CA_BIND 641 649 POTENTIAL.
FT SITE 1168 1172 GPCR MOTIF.
FT DISULFID 82 92 BY SIMILARITY.
FT DISULFID 687 696 BY SIMILARITY.
FT DISULFID 702 755 BY SIMILARITY.
FT DISULFID 807 813 BY SIMILARITY.
FT DISULFID 877 885 BY SIMILARITY.
FT DISULFID 1029 1062 BY SIMILARITY.
FT DISULFID 1066 1073 BY SIMILARITY.
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 459 459 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 779 779 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 839 839 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 882 882 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 907 907 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1007 1007 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1084 1084 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1103 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1180 AA; 130808 MW; 855DA2B502362EB4 CRC64;

Query Match 1.1%; Score 13; DB 1; Length 1180;
Best Local Similarity 100.0%; Pred. No. 0.00028;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RT human integrin alpha11 subunit (ITGA11).";
RL Genomics 60:179-187(1999).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395147; PubMed=10464311.
RA Velling T., Kusche-Gullberg M., Sejersen T., Gullberg D.,
RT "CDNA Cloning and Chromosomal Localization of Human alpha(11)
RT integrin. A collagen-binding, i domain-containing, beta(1)-associated
RT integrin alpha-chain present in muscle tissues.";
RL J. Biol. Chem. 274:25735-25742(1999).
[3]
RP SEQUENCE OF 954-1188 FROM N.A.
RT Tissue=Fibroblast;
RA Andreu N., Escivill X., Escarceller M., Sunmy L.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: INTEGRIN ALPHA-11/BETA-1 IS A RECEPTOR FOR COLLAGEN.
CC -! SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-11
CC -! ASSOCIATES WITH BETA-1.
CC -! SUBCELLULAR LOCATION: Type I membrane protein.
CC -! TISSUE SPECIFICITY: ACCORDING REF.2 HIGHEST LEVELS IN UTERUS AND
CC HEART. INTERMEDIATE LEVELS IN SKELETAL MUSCLE AND INTERMEDIATE TO
CC LOW LEVELS IN PANCREAS, KIDNEY AND PLACENTA. ACCORDING REF.1 ALSO
CC FOUND IN BRAIN, COLON, LUNG, SMALL INTESTINE, STOMACH, TESTIS,
CC SALIVARY GLANDS, THYROID GLANDS AND PROSTATE. VERY LOW LEVELS IN
CC PERIPHERAL BLOOD LYMPHOCYTES, FETAL BRAIN AND FETAL LIVER.
CC -! DEVELOPMENTAL STAGE: STRONGLY UP-REGULATED IN DIFFERENTIATING
CC FETAL MUSCLE CELLS (IN VITRO).
CC -! DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -! SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -! SIMILARITY: CONTAINS 1 VWFA DOMAIN.
CC -! SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AF109681; AAF01258.1; -
DR EMBL; AF137378; AAD51919.2; -
DR EMBL; AL359064; CAB94392.1; -
DR HSSP; P17301; 1A0X
DR Genew; HGNC:6136; ITGA11.
DR MIM; 604789; -
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; PR01185; INTEGRINA.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 5.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; FALSE_NEG.
DR PROSITE; PS02034; VWFA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium; Magnesium; Polymorphism.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 1189 INTEGRIN_ALPHA-11.
FT DOMAIN 23 1142 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1143 1165 POTENTIAL.
FT DOMAIN 1166 1189 CYTOPLASMIC (POTENTIAL).
FT REPEAT 38 94 FG-GAP 1.
FT REPEAT 102 163 FG-GAP 2.
FT REPEAT 167 345 VWFA.
FT REPEAT 359 420 FG-GAP 3.
FT REPEAT 422 475 FG-GAP 4.
FT REPEAT 477 537 FG-GAP 5.
FT REPEAT 539 598 FG-GAP 6.
FT REPEAT 601 653 FG-GAP 7.

```

```

FT DOMAIN 1154 1162 POLY-LEU.
FT DOMAIN 1174 1177 POLY-ARG.
FT CA_BIND 488 496 POTENTIAL.
FT CA_BIND 551 559 POTENTIAL.
FT CA_BIND 613 621 POTENTIAL.
FT DISULFID 76 83 BY SIMILARITY.
FT DISULFID 121 139 POTENTIAL.
FT DISULFID 129 159 POTENTIAL.
FT DISULFID 659 668 BY SIMILARITY.
FT DISULFID 674 729 BY SIMILARITY.
FT DISULFID 781 787 BY SIMILARITY.
FT DISULFID 881 893 BY SIMILARITY.
FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 642 642 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 694 694 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1032 1032 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 433 433 V -> M.
FT VARIANT 524 524 /FTID=VAR_009899.
FT VARIANT 524 524 R -> L.
FT VARIANT 972 972 /FTID=VAR_009890.
FT VARIANT 972 972 L -> P.
FT VARIANT 1003 1003 /FTID=VAR_009891.
FT VARIANT 1003 1003 I -> M.
FT VARIANT 1030 1030 /FTID=VAR_009892.
FT VARIANT 1030 1030 MISSING.
FT VARIANT 1094 1094 /FTID=VAR_009893.
FT VARIANT 1094 1094 L -> V.
FT VARIANT 1094 1094 /FTID=VAR_009894.
SQ SEQUENCE 1189 AA; 133609 MW; 60303C08A4AACD52 CRC64;

Query Match 1.1%; Score 13; DB 1; Length 1189;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RL Nature 397:176-180(1999).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE CSTA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001537; AAD06677.1; -.
CC InterPro: IPR003706; CstA.
DR Pfam: PF02554; CstA; 1.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 29 49 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 190 210 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 440 460 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 513 533 POTENTIAL.
FT TRANSMEM 546 566 POTENTIAL.
FT TRANSMEM 574 594 POTENTIAL.
FT TRANSMEM 642 662 POTENTIAL.
SQ SEQUENCE 687 AA; 74151 MW; A8BA383EDED8807 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 687;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 169 VIVLDGNSIYPM 181
DB 166 VIVLDGNSIYPM 178

RESULT 6
CSTA_HELPJ STANDARD; PRT; 687 AA.
ID CSTA_HELPJ STANDARD; PRT; 687 AA.
AC Q9ZK47;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon starvation protein A homolog.
GN CSTA OR HPI1095.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGongle B.L., Carmel G.,
RA Tumano P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
```

```

QY 1123 LMIILGSLVGG 1133
DB 118 LMIILGSLVGG 128

RESULT 7
CSTA_HELPJ STANDARD; PRT; 687 AA.
ID CSTA_HELPJ STANDARD; PRT; 687 AA.
AC P56190;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon starvation protein A homolog.
GN CSTA OR HPI168.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirnes E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weiman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
```

RL Nature 388:539-547(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE CSTA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB000622; AAD08212.1; -.
 CC TIGR; HP1168; -.
 DR InterPro: IPR003706; Csta.
 DR Pfam: PF02554; Csta; 1.
 KM Transmembrane; Inner membrane; Complete proteome.
 FT TRANSMEM 6 26
 FT TRANSMEM 29 49 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 190 210 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 256 276 POTENTIAL.
 FT TRANSMEM 285 305 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 440 460 POTENTIAL.
 FT TRANSMEM 462 482 POTENTIAL.
 FT TRANSMEM 513 533 POTENTIAL.
 FT TRANSMEM 546 566 POTENTIAL.
 FT TRANSMEM 574 594 POTENTIAL.
 FT TRANSMEM 642 662 POTENTIAL.
 SQ SEQUENCE 687 AA; 74204 MW; C525D97653E69C3D CRC64;
 Query Match 0.9%; Score 11; DB 1; Length 687;
 Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1123 LMIIGSVLGG 1133
 Db 118 LMIIGSVLGG 128
 RESULT 8
 ID ITA2_BOVIN STANDARD; PRT; 1170 AA.
 AC P53710;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (Platelet membrane glycoprotein Ia) (GP1a)
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b) (Fragment).
 GN ITGA2.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OK NCBI_TaxId=9913;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94193647; PubMed=7511592;
 RA Kamata T, Puzon W, Takada Y;
 RT "Identification of putative ligand binding sites within I domain of
 RT integrin alpha 2 beta 1 (VLA-2, CD49b/CD29).";
 RL J. Biol. Chem. 269:9659-9663(1994).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ.
 CC COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
 CC RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN
 CC COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
 CC CELLS TO COLLAGEN, MODULATION OF COLLAGEN AND COLLAGENASE GENE

CC EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L25886; AAB59255.1; -.
 CC HSRP; P17301; IAOX.
 DR InterPro: IPR000413; Integrin_alpha.
 DR InterPro: IPR02035; VWFA.
 DR Pfam; PF00092; vwa; 1.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR SMART; SMO0191; Int alpha; 5.
 DR SMART; SMO0327; VWFA_1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Platelet; Signal; Repeat; Polymorphism; Calcium; Magnesium.
 FT NON TER 1
 FT SIGNAL 1
 FT CHAIN 1 18 POTENTIAL.
 FT DOMAIN 19 1170 INTEGRIN ALPHA-2.
 FT TRANSMEM 1122 1143 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1144 1170 POTENTIAL.
 FT REPEAT 34 92 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 2 367 FG-GAP 1.
 FT REPEAT 3 367 FG-GAP 2.
 FT REPEAT 4 367 VWFA.
 FT REPEAT 5 423 FG-GAP 3.
 FT REPEAT 6 423 FG-GAP 4.
 FT REPEAT 7 423 FG-GAP 5.
 FT REPEAT 8 423 FG-GAP 6.
 FT CA_BIND 488 496 FG-GAP 7.
 FT CA_BIND 552 560 POTENTIAL.
 FT CA_BIND 616 624 POTENTIAL.
 FT SITE 472 474 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1146 1150 GEFPR MOTIF.
 FT DISULFID 72 81 BY SIMILARITY.
 FT DISULFID 669 726 BY SIMILARITY.
 FT DISULFID 778 784 BY SIMILARITY.
 FT DISULFID 854 865 BY SIMILARITY.
 FT DISULFID 1008 1039 BY SIMILARITY.
 FT DISULFID 1044 1049 BY SIMILARITY.
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 332 332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 688 688 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 748 748 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1063 1063 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1070 1070 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 580 580 G -> V.
 FT VARIANT 588 588 R -> K.
 FT VARIANT 725 725 R -> S.
 SQ SEQUENCE 1170 AA; 128929 MW; EECEP1C5F2448FB1 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 0.029;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
 DB 268 VVVTGESHHDG 278

RESULT 9
 ITA2_MOUSE STANDARD; PRT; 1178 AA.

AC 062469; Q62163;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
 GN ITGA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RX SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=94363406; PubMed=8081889;
 RA Edelman J.M., Chan B.M., Uniyal S., Onodera H., Wang D.Z.,
 RA Damjanovich L., Latzer D.B., Finberg R.W., Bergelson J.M.,
 RA "The mouse VLA-2 homologue supports collagen and laminin adhesion but
 RT not virus binding.";
 RL Cell Adhes. Commun. 2:131-143(1994).
 RP [2]
 RP SEQUENCE OF 450-1178 FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=94355691; PubMed=7521231;
 RA Wu J.E., Santoro S.A.;
 RA "Complex patterns of expression suggest extensive roles for the alpha
 RT 2 beta 1 integrin in murine development.";
 RL Dev. Dyn. 199:392-314(1994).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A COLLAGEN RECEPTOR, BEING
 CC RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER CELLS TO
 CC COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE EXPRESSION,
 CC FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 CC EXTRACELLULAR MATRIX. IT IS ALSO A RECEPTOR FOR LAMININS, COLLAGEN
 CC C-PROPEPTIDES AND E-CADHERIN. MICE HOMOZYGOUS FOR A NULL MUTATION
 CC IN THE ALPHA-2 DIE VERY EARLY IN EMBRYOGENESIS.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 CC ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z29987; CAA82877.1; -;
 CC EMBL; X75427; CAA53178.1; -;
 CC HSSP; P17301; IAOX.
 CC MGD; MGI:96600; Itga2.
 CC InterPro; IPR000413; Integrin_alpha.
 CC InterPro; IPR002035; VWF_A.
 CC Pfam; PF000692; vwa; 1.
 CC Pfam; PF00357; Integrin_A; 1.
 CC Pfam; PF01839; FG-GAP; 5.

DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWF_1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWFA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Platelet; Signal; Repeat; Calcium; Magnesium.
 FT SIGNAL 1 26
 FT CHAIN 27 1178
 FT DOMAIN 27 1129
 FT TRANSMEM 1130 1151
 FT DOMAIN 1152 1178
 FT REPEAT 42 100
 FT REPEAT ? ?
 FT REPEAT 185 375
 FT REPEAT ? ?
 FT REPEAT 431 483
 FT REPEAT 485 546
 FT REPEAT 548 607
 FT REPEAT 612 664
 FT CA_BIND 496 504
 FT CA_BIND 560 568
 FT CA_BIND 624 632
 FT SITE 480 482
 FT SITE 1154 1158
 FT DISULFID 80 89
 FT DISULFID 677 734
 FT DISULFID 786 792
 FT DISULFID 862 873
 FT DISULFID 1016 1047
 FT DISULFID 1052 1057
 FT CARBOHYD 102 102
 FT CARBOHYD 109 109
 FT CARBOHYD 429 429
 FT CARBOHYD 457 457
 FT CARBOHYD 472 472
 FT CARBOHYD 696 696
 FT CARBOHYD 1054 1054
 FT CARBOHYD 1071 1071
 FT CARBOHYD 1078 1078
 SQ SEQUENCE 1178 AA; 128926 MW; 1F194B9C0240F465 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 1178;
 Best Local Similarity 100.0%; Pred. No. 0.029;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
 DB 276 VVVTGESHHDG 286

RESULT 10
 ITA2_HUMAN STANDARD; PRT; 1181 AA.

AC P17301;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-2 precursor (platelet membrane glycoprotein Ia) (GPIa)
 DE (Collagen receptor) (VLA-2 alpha chain) (CD49b).
 GN ITGA2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 RX SEQUENCE FROM N.A. AND SEQUENCE OF 30-44.
 RC TISSUE=Endothelial cells;
 RX MEDLINE=89308879; PubMed=2545729;
 RA Takada Y., Hemler M.E.;
 RA "The primary structure of the VLA-2/collagen receptor alpha 2 subunit
 RT (platelet GPIa): homology to other integrins and the presence of a
 RT possible collagen-binding domain.";

RL J. Cell Biol. 109:397-407(1989).
 RP (2)
 RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 168-368.
 RX MEDLINE=98019223; PubMed=9353312;
 RA Emsley J., King S.L., Bergelson J.M., Liddington R.C.;
 RT "Crystal structure of the I domain from integrin alpha2beta1";
 RL J. Biol. Chem. 272:28512-28517(1997).
 RN (3)
 RP VARIANT HPA-5 (BR).
 RX MEDLINE=94043762; PubMed=7901236;
 RA Santoso S., Kalb R., Walke M., Kiefel V., Mueller-Eckhardt C.,
 RA Newman P.J.;
 RT "The human platelet alpha2beta1 integrin (alpha2) and beta2 are associated with a
 single amino acid polymorphism on glycoprotein Ia (integrin subunit
 alpha 2).";
 RL J. Clin. Invest. 92:2427-2432(1993).
 RN (4)
 RP VARIANT GLU-534.
 RX MEDLINE=20206009; PubMed=10744142; Haberbosch W., Santoso S.;
 RA Kroll H., Gardemann A., Fechter A.,
 RT "The impact of the glycoprotein Ia collagen receptor subunit A1648G
 gene polymorphism on coronary artery disease and acute myocardial
 infarction.";
 RL Thromb. Haemost. 83:392-396(2000).
 CC -1- FUNCTION: INTEGRIN ALPHA-2/BETA-1 IS A RECEPTOR FOR LAMININ,
 COLLAGEN, COLLAGEN C-PROPEPTIDES, FIBRONECTIN AND E-CADHERIN. IT
 RECOGNIZES THE PROLINE-HYDROXYLATED SEQUENCE G-F-P-G-E-R IN
 COLLAGEN. IT IS RESPONSIBLE FOR ADHESION OF PLATELETS AND OTHER
 CELLS TO COLLAGENS, MODULATION OF COLLAGEN AND COLLAGENASE GENE
 EXPRESSION, FORCE GENERATION AND ORGANIZATION OF NEWLY SYNTHESIZED
 EXTRACELLULAR MATRIX.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-2
 ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INTEGR) IS A VFMA DOMAIN. INTEGRINS
 WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- POLYMORPHISM: POSITION 534 IS ASSOCIATED WITH PLATELET-SPECIFIC
 ALLOANTIGEN HPA-5 (BR). HPA-5A/BR(A) HAS LYS-534 AND HPA-5B/BR(B)
 HAS GLU-534. HPA-5B IS INVOLVED IN NEONATAL ALLOIMMUNE
 THROMBOCYTOPENIA (NMIT OR NATP). THE K534E POLYMORPHISM MAY PLAY A
 ROLE IN CORONARY ARTERY DISEASE (CAD).
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD49b entry;
 WWW="http://www.ncbi.nlm.nih.gov/ncbi/cd/cd49b.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X17033; CAA34894.1; -
 DR PIR: A33998; A33998.
 DR PDB: 1A0X; 25-NOV-98.
 DR Genew; HGNC:6137; ITGA2.
 DR MIM; 192974; -
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VFMA_A.
 DR Pfam; PF00092; vma; 1.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PR00453; VFMADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; vma; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VFMA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Platelet; Signal; Repeat; Calcium; Magnesium; Polymorphism;
 KW 3D-structure.

FT	SIGNAL	1	29	
FT	CHAIN	30	1181	INTEGRIN ALPHA-2.
FT	DOMAIN	30	1132	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1133	1154	POTENTIAL.
FT	DOMAIN	1155	1181	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	45	103	FG-GAP 1.
FT	REPEAT	?	?	FG-GAP 2.
FT	DOMAIN	188	378	VFMA.
FT	REPEAT	378	433	FG-GAP 3.
FT	REPEAT	434	486	FG-GAP 4.
FT	REPEAT	488	549	FG-GAP 5.
FT	REPEAT	551	610	FG-GAP 6.
FT	REPEAT	615	667	FG-GAP 7.
FT	CA_BIND	499	507	POTENTIAL.
FT	CA_BIND	563	571	POTENTIAL.
FT	CA_BIND	627	635	POTENTIAL.
FT	SITE	1157	1161	GFPR MOTIF.
FT	DISULFID	83	92	BY SIMILARITY.
FT	DISULFID	680	737	BY SIMILARITY.
FT	DISULFID	789	795	BY SIMILARITY.
FT	DISULFID	865	876	BY SIMILARITY.
FT	DISULFID	1019	1050	BY SIMILARITY.
FT	DISULFID	1055	1060	BY SIMILARITY.
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	343	343	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	432	432	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	460	460	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	475	475	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	699	699	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1057	1057	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1074	1074	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1081	1081	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	534	534	K -> E (IN ALLOANTIGEN HPA-5B).
FT	SEQUENCE	1181 AA;	129295 MW;	7E1B7ED968A94070 CRC64;
SO	Query Match	0.9%;	Score 11;	DB 1; Length 1181;
	Best Local Similarity	100.0%;	Pred. No. 0.029;	
	Matches 11;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	272	VVVTGDSHSDG	282	
DB	279	VVVTGDSHSDG	289	
RESULT 11				
ITAE_HUMAN	STANDARD;	PRT;	1179 AA.	
ID	P38570; Q9NZU9;			
AC	01-OCT-1994 (Rel. 30. Created)			
DT	30-MAY-2000 (Rel. 39, last sequence update)			
DT	15-JUN-2002 (Rel. 41, last annotation update)			
DE	Integrin alpha-E precursor (Mucosal lymphocyte-1 antigen) (HML-1 antigen) (CD103 antigen) (Integrin alpha-IEH).			
GN	ITGA2			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	(1)			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 19-38 AND 179-188.			
RC	TISSUE=Lymphocytes, and Leukemia;			
RX	MEDLINE=94164962; PubMed=8119947;			
RA	Shaw S.K., Cepek K.L., Murphy E.A., Russell G.J., Brenner M.B.,			
RT	"Molecular cloning of the human mucosal lymphocyte integrin alpha E			
RT	subunit. Unusual structure and restricted RNA distribution.";			
RL	J. Biol. Chem. 269:6016-6025(1994).			
RN	(2)			
RP	REVISIONS TO 88-114.			
RA	Parker C.M.;			
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.			

```

RN [3]
RP SEQUENCE OF 53-1179 FROM N.A.
RC TISSUE=Fetal kidney;
RA MEDLINE=20138496; PubMed=10673275;
RA Touchman J.W., Ankster Y., Dietrich N.L., Maduro V.V., McDowell G.,
RA Shostersuk V., Bouffard G.G., Beckstrom-Sternberg S.M., Gahl W.A.,
RA Green E.D.;
RT "The genomic region encompassing the nephropathic cystinosis gene
RT (CTNS): complete sequencing of a 200-kb segment and discovery of a
RT novel gene within the common cystinosis-causing deletion.";
RL Genome Res. 10:165-173(2000).
RN [4]
RP MUTAGENESIS OF ASP-109 AND PHE-316.
RX MEDLINE=20400502; PubMed=10837471;
RX Hsiangins J.M.H., Cernadas M., Tan K., Irie A., Wang J.-H., Takada Y.,
RA Brenner M.B.;
RT "The role of alpha and beta chains in ligand recognition by beta 7
RT integrins.";
RL J. Biol. Chem. 275:25652-25664(2000).
CC -1- FUNCTION: INTEGRIN ALPHA-E/BETA-7 IS A RECEPTOR FOR E-CADHERIN. IT
CC MEDIATES ADHESION OF INTRA-EPITHELIAL T-LYMPHOCYTES TO EPITHELIAL
CC CELL MONOLAYERS.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAINS LINKED BY A
CC DISULFIDE BOND. ALPHA-E ASSOCIATES WITH BETA-7.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED ON A SUBCLASS OF T-LYMPHOCYTES KNOWN
CC AS INTRA-EPITHELIAL LYMPHOCYTES WHICH ARE LOCATED BETWEEN MUCOSAL
CC EPITHELIAL CELLS.
CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A WFPA DOMAIN. INTEGRINS
CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 1 WFPA DOMAIN.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
CC -1- DATABASE: NAME=PRO; NOTE=CD guide CD103 entry;
CC WWW=ftp://www.ncbi.nlm.nih.gov/prov/cd/cd103.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L25851; AAB59359.2; -
DR EMBL: AF168787; AAF43107.1; -
DR HSSP: P11215; 1A8X.
DR GeneW; HGNC:6147; ITGAE.
DR MIM: 604682; -
DR InterPro: IPR000413; Integrin_alpha.
DR InterPro: IPR002035; WMF_A.
DR Pfam: PF00092; wmf_1.
DR Pfam: PF00357; Integrin_A_1.
DR Pfam: PF01839; FG-GAP_4.
DR PRINTS: PRO1185; INTEGRINA.
DR PRINTS: PRO0453; WMFADOMAIN.
DR SMART: SM00191; Int_alpha_4.
DR SMART: SM00327; WMA_1.
DR PROSITE: PS00242; INTEGRIN_ALPHA_1.
DR PROSITE: PS00234; WMFA_1.
DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Polymorphism; Magnesium; Calcium.
FT SIGNAL 1 18
FT CHAIN 19 1179 INTEGRIN ALPHA-E.
FT CHAIN 19 1177 INTEGRIN ALPHA-E LIGHT CHAIN.
FT CHAIN 179 1177 INTEGRIN ALPHA-E HEAVY CHAIN.
FT DOMAIN 19 1124 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1125 1147 POTENTIAL.
FT DOMAIN 1148 1179 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 181 198 GLU-RICH (ACIDIC).
FT REPEAT ? ? FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.

```

```

FT DOMAIN 145 199 X-DOMAIN (EXTRA DOMAIN).
FT DOMAIN 200 391 WFPA.
FT REPEAT 401 456 FG-GAP 3.
FT REPEAT 457 506 FG-GAP 4.
FT REPEAT 510 571 FG-GAP 5.
FT REPEAT 573 658 FG-GAP 6.
FT REPEAT 641 693 FG-GAP 7.
FT CA_BIND 522 530 POTENTIAL.
FT CA_BIND 586 594 POTENTIAL.
FT CA_BIND 654 662 POTENTIAL.
FT SITE 1150 1154 GPRK MOTIF.
FT DISULFID 70 79 BY SIMILARITY.
FT DISULFID 126 159 BY SIMILARITY.
FT DISULFID 706 762 BY SIMILARITY.
FT DISULFID 823 829 BY SIMILARITY.
FT DISULFID 893 907 BY SIMILARITY.
FT DISULFID 1008 1033 BY SIMILARITY.
FT DISULFID 1041 1057 BY SIMILARITY.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 934 934 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 954 954 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1065 1065 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 360 360 D -> E.
FT VARIANT 1041 1041 /FTID=VAR_008884.
FT VARIANT 1041 1041 C -> S.
FT MUTAGEN 208 208 /FTID=VAR_008885.
FT MUTAGEN 316 316 D -> A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 477 477 F -> A: LOSS OF E-CADHERIN BINDING.
FT CONFLICT 482 482 V -> I (IN REF. 3).
FT CONFLICT 950 950 Q -> R (IN REF. 3).
FT CONFLICT 950 950 R -> W (IN REF. 3).
FT CONFLICT 1019 1019 A -> V (IN REF. 3).
SQ SEQUENCE 1179 AA; 13008 MW; E558902EDP9D95E1 CRC64;

Query Match Score 10; DB 1; Length 1179;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSELCIP 492
DB 511 GSYFSELCIP 520

RESULT 12
ID ITAL_MOUSE STANDARD; PRT; 1163 AA.
AC P24063;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-L precursor (Leukocyte adhesion glycoprotein LFA-1
DE alpha chain) (Leukocyte function associated molecule 1, alpha chain)
DE (CD11a).
GN ITGAL OR LFA-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91268576; PubMed=2051027;
RA Kaufmann Y., Tseng E., Springer T.A.;
RT "Cloning of the murine lymphocyte function-associated molecule-1
RT alpha-subunit and its expression in COS cells.";
RL J. Immunol. 147:369-374(1991).

```


[illegible]

Query Match	Best Local Similarity	0.84; Score 9;	DB 1;	Length 1163;
Matches	9;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
Qy	566 DVAVGAPLE 574			
Db	536 DVAVGAPLE 544			
RESULT 13				
ITAL_HUMAN	STANDARD;		PRT;	1170 AA.
ID	P20701; 043746;			
AC	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Integrin alpha-L precursor (leukocyte adhesion glycoprotein LFA-1 alpha chain) (leukocyte function associated molecule 1, alpha chain) (CD11a).			
DE	CD11a).			
GN	ITGAL OR CD11A.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.			
RX	MEDLINE=89139587; Pubmed=2537322;			
RA	larsen R.S., Corbi A.L., Berman L., Springer T.;			
RT	"Primary structure of the leukocyte function-associated molecule-1 alpha subunit: an integrin with an embedded domain defining a protein superfamily.";			
RL	J. Cell Biol. 108:703-712(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 2).			
RX	MEDLINE=99425270; Pubmed=10493829;			
RA	Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,			
RA	Fuhrmann J., Mason T., Crosby M.L., Barneshead M., Cronin L.,			
RA	Delacates Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,			
RA	Diehlert E.E., Harris P.C., Venter J.C., Adams M.D.;			
RT	"Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q.";			
RL	Genomics 60:295-308(1999).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 153-335, AND REVISION TO 214			
RX	MEDLINE=9603067; Pubmed=7479767;			
RA	Qu A., Leahy D.J.;			
RT	"Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, alpha			
RL	L beta 2) integrin.";			
RN	Proc. Natl. Acad. Sci. U.S.A. 92:10277-10281(1995).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 153-335.			
RX	MEDLINE=9639682; Pubmed=8805579;			
RA	Qu A., Leahy D.J.;			
RT	"The role of the divalent cation in the structure of the I domain			
RL	from the CD11a/CD18 integrin.";			
	Structure 4:931-942(1996).			

[5] X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 153-334.
 MEDLINE=99425288; PubMed=10493852;
 RA Kallen J, Welzenbach K, Ramage P, Geyl D, Krawacki R, Legge G,
 RA Cottens S, Weitz-Schmidt G, Hommel U;
 RT "Structural basis for LFA-1 inhibition upon Iovastatin binding to the
 RT CD11a 1-domain.";
 RT J. Mol. Biol. 292:1-9(1999).
 CC -1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
 CC ICAM3 AND ICAM4. IT IS INVOLVED IN A VARIETY OF IMMUNE PHENOMENA
 CC INCLUDING LEUKOCYTE-ENDOTHELIAL CELL INTERACTION, CYTOTOXIC T-CELL
 CC MEDIATED KILLING, AND ANTIBODY DEPENDENT KILLING BY GRANULOCYTES
 CC AND MONOCYTES.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-L
 CC ASSOCIATES WITH BETA-2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: LEUKOCYTES.
 CC -1- DOMAIN: THE INTEGRIN 1-DOMAIN (INSERT) IS A VWFA DOMAIN. INTEGRINS
 CC WITH 1-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11a entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd11a.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, Y00796; CAA68747.1; --
 DR EMBL, AC002310; AAC31672.1; --
 DR PIR, S03308; S03308.
 DR PDB, 1LFA; 29-JAN-96.
 DR PDB, 1ZON; 07-DEC-96.
 DR PDB, 1ZOO; 07-DEC-96.
 DR PDB, 1ZOP; 07-DEC-96.
 DR PDB, 1COP; 07-AUG-00.
 DR Genew; HGNC:6148; ITGAL.
 DR MIM, 153370; --
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWFA.
 DR Pfam; PF00092; vwa; 1.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF01839; FG-GAP; 5.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWFA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS5024; VWFA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KM Signal; 3D-structure; Magnesium; Calcium; Repeat;
 KM Alternative splicing
 FT SIGNAL 1 25 INTEGRIN ALPHA-L.
 FT CHAIN 26 1170 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 26 1088 POTENTIAL.
 FT TRANSMEM 1089 1112 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1113 1170
 FT REPEAT 42 91 FG-GAP 1.
 FT REPEAT 92 143 FG-GAP 2.
 FT DOMAIN 170 349 VWFA.
 FT REPEAT 170 349 FG-GAP 3.
 FT REPEAT 401 455 FG-GAP 4.
 FT REPEAT 457 516 FG-GAP 5.
 FT REPEAT 518 575 FG-GAP 6.
 FT REPEAT 578 630 FG-GAP 7.
 FT CA_BIND 468 476 POTENTIAL.

FT	CA_BIND	530	538	POTENTIAL.
FT	CA_BIND	530	538	POTENTIAL.
FT	CA_BIND	530	538	POTENTIAL.
FT	SITE	1115	1119	GFPR MOTIF.
FT	DISULFID	73	80	BY SIMILARITY.
FT	DISULFID	111	129	BY SIMILARITY.
FT	DISULFID	653	707	BY SIMILARITY.
FT	DISULFID	771	777	BY SIMILARITY.
FT	DISULFID	845	861	BY SIMILARITY.
FT	DISULFID	998	1013	BY SIMILARITY.
FT	DISULFID	1021	1052	BY SIMILARITY.
FT	CARBOHYD	65	65	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	89	89	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	649	649	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	670	670	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	726	726	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	730	730	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	862	862	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	885	885	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	897	897	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1060	1060	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1071	1071	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	954	954	O -> OGCHGVEMQTSQIICRPGAGDEHTVGAEGELPC PWGSEARNDIRAGPCR (IN ISOFORM 2). R -> W (IN REF. 1 AND 2). Y -> I (IN REF. 2).
FT	CONFLICT	214	214	
FT	CONFLICT	660	660	
FT	SEQUENCE	1170 AA;	128819 MW;	39A7AF92EF286FC0 CRC64;

Query Match 0.8%; Score 9; DB 1; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 566 DVAVGAPLE 574
 538 DVAVGAPLE 546

RESULT 14
 CD24_MOUSE STANDARD; PRT; 76 AA.
 AC P24807; P26691;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Signal transducer CD24 precursor (M1/69-J11D heat stable antigen)
 DE (HSA) (Nectadidin) (LY-52) (X62 heat stable antigen) (R13-AG).
 GN CD24 OR CD24A.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90361906; PubMed=2118158;
 RA Kay R., Takei F., Humphries R.K.;
 RT "Expression cloning of a cDNA encoding M1/69-J11D heat-stable
 RT antigens.";
 RT J. Immunol. 145:1952-1959(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA, and Swiss albino X BALB/c; TISSUE=Spleen;
 RX MEDLINE=91209380; PubMed=2019286;
 RA Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen P.J.;
 RT "The genes for a mouse hematopoietic differentiation marker called
 RT the heat-stable antigen.";
 RT Eur. J. Immunol. 21:1039-1046(1991).
 RN [3]
 RP REVISIONS.
 RA Nielsen P.J.;
 RL Submitted (Jul-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Spleen;

RX MEDLINE=94043127; PubMed=8226859;
 RA Wenger R.H., Rochelle J.M., Seldin M.F., Koehler G., Nielsen P.U.;
 RT "The heat stable antigen (mouse CD24) gene is differentially
 RT regulated but has a housekeeping promoter.";
 RL J. Biol. Chem. 268:23345-23352(1993).
 RN [5]
 RP SEQUENCE OF 27-53.
 RC STRAIN=C57BL/6;
 RX MEDLINE=92412120; PubMed=1530634;
 RA Hiteamco Y., Nakano A., Ohnishi H., Hamada F., Saheki S.,
 RA Takeuchi N.;
 RT "Purification of the murine heat-stable antigen from erythrocytes";
 RL Biochem. Biophys. Res. Commun. 187:773-777(1992).
 CC -1- FUNCTION: MAY HAVE A SPECIFIC ROLE TO PLAY IN EARLY THYMOCYTE
 CC DEVELOPMENT.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: IN LYMPHOID, MYELOID, AND ERYTHROID CELLS.
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, MS8661; AAA39481.1; -
 DR EMBL, X56469; CAA39841.1; -
 DR EMBL, X72910; CAA51415.1; -
 DR EMBL, X53825; CAA37822.1; -
 DR PIR, S43709; S43709.
 DR PIR, A43537; A43537.
 DR PIR, S32240; S32240.
 DR PIR, S15784; S15784.
 DR PIR, S33129; S33129.
 DR PIR, S19111; S19111.
 DR MGI, 88323; Cd24a.
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Antigen.
 FT SIGNAL 1 26
 FT CHAIN 1 26
 FT PROPEP 54 53
 FT CARBOHYD 27 76
 FT CARBOHYD 30 27
 FT CARBOHYD 39 30
 FT CARBOHYD 48 39
 FT LIPID 53 48
 FT SEQUENCE 76 AA; 7797 MW; 6853F12183625EB CRC64;
 SQ
 Query Match 0.7%; Score 8; DB 1; Length 76;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1133 GILLIAL 1140
 Db 12 GILLIAL 19
 RESULT 15
 ID CD24 RAT STANDARD; PRT; 76 AA.
 AC 007450;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Signal transducer CD24 precursor (Heat stable antigen) (HSA)
 DE (Nectadrin).
 GN CD24A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OK NCBI_TaxID=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=wislar; TISSUE=Embryonic brain;
 RX MEDLINE=9412434; PubMed=8292828;
 RA Shimasawa T., Akashi T., Sakamoto K., Takahashi H., Maruyama N.,
 RA Hirokawa K.;
 RT "Gene expression of CD24 core peptide molecule in developing brain
 RT and developing non-neural tissues.";
 RL Dev. Dyn. 198:1-13(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fischer;
 RX MEDLINE=97157759; PubMed=9004038;
 RA Magnalio T.A., Barandon Y.;
 RT "CD24 (heat stable antigen, nectadrin), a novel keratinocyte
 RT differentiation marker, is preferentially expressed in areas of the
 RT hair follicle containing the colony-forming cells";
 RL J. Cell Sci. 109:3035-3045(1996).
 CC -1- FUNCTION: MAY HAVE A PIVOTAL ROLE IN CELL DIFFERENTIATION. THE
 CC TRIGGERING MECHANISM OF SIGNAL TRANSDUCTION MAY BE DUE TO THE
 CC INTERACTIONS OF DIFFERENTIATING CELLS WITH THE MATRIX SUBSTRATE
 CC VIA THE CARBOHYDRATE STRUCTURE OF THE MOLECULE. IN THIS WAY, THE
 CC SIGNAL TRANSDUCER CAN PLAY VERY DIFFERENT ROLES IN DIFFERENT CELL
 CC TYPES AS A DIRECT CONSEQUENCE OF ITS GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE CENTRAL NERVOUS SYSTEM, IN
 CC POSTMITOTIC CELLS OF SPINAL CORD, HINDBRAIN, MIDBRAIN AND
 CC FOREBRAIN. EXPRESSED IN EPITHELIUM DURING THE DEVELOPMENT OF NON-
 CC NEURAL TISSUES. EXPRESSED IN TOOTH DEVELOPMENT, SPECIFICALLY IN
 CC MESENCHYMAL CELLS DIFFERENTIATING INTO ODONTOBLAST IN DENTAL
 CC PAPILLA, AS WELL AS IN THE DEVELOPING EYE AND HAIR FOLLICLE.
 CC -1- DEVELOPMENTAL STAGE: DETECTED IN PRIMITIVE ECTODERM, MESODERM AND
 CC VENTRAL ENDODERM, DOWN-REGULATED WHEN ORGANOGENESIS IS COMPLETED.
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY). THE CARBOHYDRATE
 CC STRUCTURE MAY BE REGULATED IN A TISSUE SPECIFIC AND DEVELOPMENTAL
 CC STAGE SPECIFIC MANNER.
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, Z11663; CA77731.1; -
 DR EMBL, U49062; AAA91470.1; -
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Differentiation.
 FT SIGNAL 1 26
 FT CHAIN 1 26
 FT PROPEP 57 76
 FT CARBOHYD 27 56
 FT CARBOHYD 37 27
 FT CARBOHYD 48 37
 FT LIPID 56 48
 FT SEQUENCE 76 AA; 7862 MW; 42846E70EC39D958 CRC64;
 SQ
 Query Match 0.7%; Score 8; DB 1; Length 76;
 Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1133 GILLIAL 1140
 Db 12 GILLIAL 19
 RESULT 16
 ID CD24 HUMAN STANDARD; PRT; 80 AA.
 AC P25063; Q16257;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Signal transducer CD24 precursor.
GN CD24 OR CD24.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91332458; PubMed=1831224;
RA Kay R., Roeten P.M., Humphries R.K.;
RT "CD24, a signal transducer modulating B cell activation responses, is
RT a very short peptide with a glycosyl phosphatidylinositol membrane
RT anchor."
RL J. Immunol. 147:1412-1416(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93007871; PubMed=1327504;
RA Jackson D., Walbel R., Weber E., Bell J., Stachel R.A.;
RT "CD24, a signal-transducing molecule expressed on human B cells, is a
RT major surface antigen on small cell lung carcinomas."
RL Cancer Res. 52:5264-5270(1992).
RN [3]
RP SEQUENCE OF 1-76 FROM N.A.
RX MEDLINE=95048364; PubMed=7959762;
RA Hough M.R., Roeten P.M., Sexton T.L., Kay R., Humphries R.K.;
RT "Mapping of CD24 and homologous sequences to multiple chromosomal
RT loci."
RL Genomics 22:154-161(1994).
CC -1- FUNCTION: MODULATES B-CELL ACTIVATION RESPONSES. SIGNALING COULD
CC BE TRIGGERED BY THE BINDING OF A LECTIN-LIKE LIGAND TO THE CD24
CC CARBOHYDRATES, AND TRANSDUCED BY THE RELEASE OF SECOND MESSAGES
CC DERIVED FROM THE GPI-ANCHOR. PROMOTES AG-DEPENDENT PROLIFERATION
CC OF B-CELLS. AND PREVENTS THEIR TERMINAL DIFFERENTIATION INTO
CC ANTIBODY-FORMING CELLS.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: B-CELLS.
CC -1- INDUCTION: EXPRESSION IS LOST WHEN PRIMARY B-CELLS ARE INDUCED TO
CC DIFFERENTIATE IN ANTIBODY-FORMING CELLS.
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD24 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd24.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M58664; AAA35665.1; -;
DR EMBL, X69397; CA449195.1; -;
DR EMBL, L33930; AAB58807.1; -;
DR EMBL, S75311; AAD14170.1; ALT_INIT.
DR Genew; HGNC:1645; CD24.
DR MIM: 600074; -;
DR KW Glycoprotein; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 59
FT PROPEP 60 80
FT CARBOHYD 36 36
FT CARBOHYD 52 52
FT LIPID 59 59
FT CONFLICT 12 12
FT CONFLICT 44 44
FT CONFLICT 57 57
FT CONFLICT 57 57
FT SEQUENCE 80 AA; 8083 MW; DB1988B6808F833F CRC64;
SO SEQUENCE

Query Match 0.7%; Score 8; DB 1; Length 80;
Best Local Similarity 100.0%; Pred.No.3.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GELLALL 1140
DB 12 GELLALL 19
RESULT 17
PRVB MERBI STANDARD; PRT; 108 AA.
AC P56503;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Parvalbumin beta.
OS Merluccius bilinearis (Silver hake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Merlucciidae;
OC Merluccius.
OX NCBI_TaxID=79698;
RN [1]
RP SEQUENCE.
RX MEDLINE=98046752; PubMed=9385642;
RA Revett S.P., King G., Shabanowitz J., Hunt D.F., Hartman T.M.,
RA Nelson D.J.;
RT "Characterization of a helix-loop-helix (EF hand) motif of silver
RT hake parvalbumin isoform B."
RL Protein Sci. 6:2397-2408(1997).
CC -1- FUNCTION: IN MUSCLE, THE CALCIUM-BINDING PROTEIN PARVALBUMIN IS
CC THOUGHT TO BE INVOLVED IN MUSCLE RELAXATION.
CC -1- MASS SPECTROMETRY: MW=11357; MW ERR=3.5; METHOD=Electrospray.
CC -1- MISCELLANEOUS: THIS PARVALBUMIN HAS AN ISOELECTRIC POINT OF 4.2.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BELONGS TO
CC THE PARVALBUMINS SUBFAMILY.
DR HSP: P02621; 1475.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 2.
DR Prodom; PD000012; EF-hand; 1.
DR PROSITE; PS00018; EF-HAND; 2.
KW Calcium-binding; Muscle protein; Repeat; Acetylation.
FT MOD RES 1 1
FT CA BIND 51 62
FT CA BIND 90 101
FT CA BIND 90 101
FT SEQUENCE 108 AA; 11317 MW; 08788BFC54002906 CRC64;
SO SEQUENCE

Query Match 0.7%; Score 8; DB 1; Length 108;
Best Local Similarity 100.0%; Pred.No.4.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 287 AALKACEA 294
DB 13 AALKACEA 20
RESULT 18
ID IL22 HUMAN STANDARD; PRT; 179 AA.
AC O9GZX6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Interleukin-22 precursor (IL-22) (IL-10-related T-cell-derived
DE inducible factor) (IL-TIF).
GN IL22 OR ILTIF
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20420346; PubMed=10954742;
RA Dumoutier L., Van Roost E., Colau D., Renaud J.-C.,
RT "Human interleukin-10-related T cell-derived inducible factor:

RT molecular cloning and functional characterization as an hepatocyte-stimulating factor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:10144-10149(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21069354; PubMed=1197690;
 RA Dumoutier L., Van Roost E., Coliau D., Ameys G., Michaux L.,
 RT Renauld J.-C.;
 RL "IL-TIF/IL-22: genomic organization and mapping of the human and mouse genes";
 RN Genes Immun. 1:488-494(2000).
 RP [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20469498; PubMed=10875937;
 RA Xie M.-H., Aggarwal S., Ho W.-H., Foster J., Zhang Z., Stinson J.,
 RT Wood W. I., Goddard A. D., Gunney A. L.;
 RL "Interleukin (IL)-22, a novel human cytokine that signals through the interferon receptor-related proteins CRF2-4 and IL-22R.";
 RN J. Biol. Chem. 275:31335-31339(2000).
 RN [4]
 RP SEQUENCE FROM N.A., AND VARIANT GUY-158.
 RA Rieder M. J., Carrington D. P., Chung M.-W., Lee K. L., Poel C. L., Yi Q.,
 RL Nickerson D. A.;
 CC Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Cytokine that contributes to the inflammatory response in vivo.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE IL-10 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ277247; CAC06085.1; -
 DR EMBL: AJ277248; CAC19409.1; -
 DR EMBL: AF279437; AAG22064.1; -
 DR EMBL: AF367519; AAK62468.1; -
 DR Genew; HGNC:14900; IL22.
 DR MIM; 605330; -
 DR InterPro; IPR000098; Interleukin_10.
 DR SMART; SM00188; IL10; 1.
 DR PROSITE; PS00520; INTERLEUKIN_10; 1.
 KM Cytokine; Glycoprotein; Signal; Polymorphism.
 FT SIGNAL 1 33
 FT CHAIN 1 179
 FT CARBOHYD 54 54
 FT CARBOHYD 68 68
 FT CARBOHYD 97 97
 FT VARIANT 158 158
 FT SEQUENCE 179 AA; 20011 MW; 3C35E64D60CF8767 CRC64;
 SQ
 Query Match 0.7%; Score 8; DB 1; Length 179;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1134 LLLALLLV 1141
 Db 21 LLLALLLV 28
 RESULT 19
 EMB2_CAVPO STANDARD; PRT; 234 AA.
 AC P35709;
 DT 01-FEB-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Eosinophil granule major basic protein 2 precursor (MBP-2).

GN MBP2.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
 OC NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Eosinophil;
 RX MEDLINE=91224343; PubMed=2026266;
 RA Aoki I., Shindoh Y., Nishida T., Nakai S., Hong Y.-M., Mio M.,
 RT Saito T., Taseka K.;
 RL "Comparison of the amino acid and nucleotide sequences between human and two guinea pig major basic proteins.";
 RL FEBS Lett. 282:56-60(1991).
 CC -1- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC REACTIONS AND INFLAMMATIONS, SINCE MBP IS CAPABLE OF RELEASING HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.
 CC HITAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF BRONCHIAL TUBES. ANTIPARASITIC AND ANTIBIOTIC.
 CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC GRANULE (CRYSTALLOID CORE).
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D0817; BAA0697.1; -
 DR PIR; S15102; S15102.
 DR HSSP; P13727; IH8U.
 DR InterPro; IPR002352; Emaior basic.
 DR InterPro; IPR001304; Lectin_C.
 DR Pfam; PF00059; Lectin_c; 1.
 DR PRINTS; PR00770; EMBJORBASICP.
 DR SMART; SM00034; CLECT; 1.
 DR PROSITE; PS00615; C-TYPE_LLECTIN_1; 1.
 DR PROSITE; PS50041; C-TYPE_LLECTIN_2; 1.
 KM Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
 KM Lectin; Multigene family; Glycoprotein.
 FT SIGNAL 1 15
 FT PROPEP 1 115
 FT CHAIN 116 234
 FT DOMAIN 133 234
 FT DISULFID 135 232
 FT DISULFID 209 224
 FT CARBOHYD 69 69
 FT SEQUENCE 234 AA; 26140 MW; 7D926A942BF5116F CRC64;
 SQ
 Query Match 0.7%; Score 8; DB 1; Length 234;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1134 LLLALLLV 1141
 Db 4 LLLALLLV 11
 RESULT 20
 PSA7 YEAST STANDARD; PRT; 254 AA.
 AC P40303;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Proteasome component PRB6 (EC 3.4.25.1) (Macropain subunit PRB6) (Proteinase YSCB subunit PRB6) (Multicatalytic endopeptidase complex subunit PRB6).
 DE PRB6 OR YOL038W.
 GN Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c;
 RX MEDLINE=95001940; PubMed=7918444;
 RA Heinemeyer W., Troendle N., Albrecht G., Wolf D.H.;
 RT "PRE5 and PRE6, the last missing genes encoding 20S proteasome
 subunits from yeast? Indication for a set of 14 different subunits in
 the eukaryotic proteasome core.";
 RL Biochemistry 33:12229-12237(1994).
 RP SEQUENCE FROM N.A.
 RA Ansoorge W., Bernes V., Rechmann S., Schwager C., Teodoru C., Voss H.,
 RA Wiemann S.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 112-133.
 RA Troendle N.;
 RL Thesis (1991), University of Stuttgart, Germany.
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 3-343.
 RX MEDLINE=97242404; PubMed=9087403;
 RA Groll M., Ditzel U., Lowe J., Stock D., Boehler M., Bartunik H.D.,
 RA Huber R.;
 RL "Structure of 20S proteasome from yeast at 2.4-A resolution.";
 RT Nature 386:463-471(1997).
 CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
 WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
 PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
 SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
 ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
 specificity.
 CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
 proteolytic pathway.
 CC -1- SUBUNIT: The yeast proteasome seems to be composed of 14 different
 subunits which form a highly ordered ring-shaped structure.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L34348; AAA34903.1; -
 DR EMBL: Z74780; AAA39040.1; -
 DR PIR: S47910; S47910.
 DR PDB: 1KVP; 15-APR-98.
 DR MEROPS: T01.974; -
 DR SGD: S0005398; PRE6.
 DR InterPro: IPR000426; Proteasome_A.
 DR InterPro: IPR001353; Proteasome_protease.
 DR Pfam: PF00227; Proteasome_1.
 DR PROSITE: PS00188; PROTEASOME_A; 1.
 KW Proteasome; Hydrolyase; Protease; 3D-structure.
 SQ SEQUENCE 254 AA; 28439 MW; 73AE65B836E618A CRC64;

Query Match 0.7%; Score 8; DB 1; Length 254;
 Best Local Similarity 100.0%; Pred. No. 8.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1109 SLEEVQT 1116
 |||||

DR 198 SLEEVQT 205
 RESULT 21
 ID HISU_CAMJ6
 AC 046125; Q9EPH3;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Histidine-binding protein precursor (HBP) (p29).
 GN HISU OR CDAC OR CU0734C.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxId=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M275;
 RX MEDLINE=9735360; PubMed=9192026;
 RA Garvis S.G., Puzon G.J., Konkel M.E.;
 RT "Cloning, sequencing, and expression of a Campylobacter jejuni
 periplasmic binding protein (p29) involved in histidine transport.";
 RL Adv. Exp. Med. Biol. 412:263-264(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=72Dz/92;
 RX MEDLINE=98055437; PubMed=9395059;
 RA Pawelec D., Kozyniek B., Popowski J., Jaguszczyn-Krynicka E.K.;
 RT "Cloning and characterization of a Campylobacter jejuni 72Dz/92 gene
 encoding a 30 kDa immunopositive protein, component of the ABC
 transport system; expression of the gene in avirulent salmonella
 typhimurium.";
 RL FEBS Immunol. Med. Microbiol. 19:137-150(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Paulwell T., Holtroyd S.,
 RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrett B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 CC -1- FUNCTION: COMPONENT OF THE HIGH-AFFINITY HISTIDINE PERMEASE. A
 BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM. THE OTHER COMPONENTS
 ARE PROTEINS HISQ, M, AND P (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Probable).
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
 PROTEIN FAMILY 3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U58045; AAC35419.1; -
 DR EMBL: Y10873; CA871823.1; -
 DR HSSP: P39182; IHS1.
 DR InterPro: IPR001311; SBP_glu_receptor.
 DR InterPro: IPR001638; SBP_bac_3.
 DR Pfam: PF00497; SBP_bac_3; 1.
 DR ProDom: PD000500; Ton_glu_receptor; 1.
 DR SMART: SM00062; PBP; 1.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 DR PROSITE: PS01039; SBP_BACTERIAL_3; 1.

KW Transport; Amino-acid transport; Signal; Membrane; Lipoprotein;
 Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 256 HISTIDINE-BINDING PROTEIN.
 FT LIPID 20 20 N-ACYL GLYCERIDE (PROBABLE).
 FT CONFLICT 4 8 FLTAIF -> ILSTA (IN REF. 3).
 FT CONFLICT 12 13 FT -> LV (IN REF. 3).
 FT CONFLICT 18 38 VACONTENNANSNENTLIT -> GACSDSKNKESNAVE
 (IN REF. 3).
 FT CONFLICT 52 55 FKOD -> YKEN (IN REF. 3).
 FT CONFLICT 66 66 I -> V (IN REF. 3).
 FT CONFLICT 76 76 E -> K (IN REF. 3).
 FT CONFLICT 93 93 S -> A (IN REF. 3).
 FT CONFLICT 228 228 D -> N (IN REF. 3).
 SQ SEQUENCE 256 AA; 28531 MW; 2E3E34DFCB92CB29 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 8; DB 1; Length 256;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 922 GTLOENTA 929
 DB 151 GTLOENTA 158
 RESULT 22
 PSAT SCHPO STANDARD; PRT; 259 AA.
 ID PSAT SCHPO
 AC 010329;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable proteasome subunit alpha type 7 (EC 3.4.25.1).
 GN SPBC106.16.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OC NCBI_Taxid=4896;
 RX NCBI_Taxid=4896;
 RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouras J., Peat N., Hayles J., Baker S., Bigham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Stammers M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Grynopreuz B.,
 Welfens I., Vansteelandt E., Rieger M., Schaefer M., Meller-Auer S.,
 Gabel C., Fuchs M., Fritz C., Holzer E., Mosel D., Hilbert H.,
 Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Medler H., Wandut R., Purrelle B.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 Daga R.R., Cruzada J.L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Shpkovski G.V., Useery D., Barrett B.G., Nurse P.;
 "The genome sequence of Schizosaccharomyces pombe";
 Nature 415:871-880 (2002).
 -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
 WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
 PHE, TR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
 SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC

CC ACTIVITY (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
 CC specificity.
 CC -!- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
 CC proteolytic pathway.
 CC -!- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
 CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AL110295; CAB53732.1; -.
 CC
 CC DR HSPD; P40303; IRP.
 CC DR MEROPS; T01.974; -.
 CC
 CC DR InterPro; IPR000426; Proteasome_A.
 CC DR InterPro; IPR001353; Proteasome.
 CC DR Pfam; PF00227; Proteasome; 1.
 CC DR PROSITE; PS00388; PROTEASOME_A; 1.
 CC KW Proteasome; Hydrolase; Protease.
 CC SEQUENCE 259 AA; 28277 MW; 6F36070FE531380 CRC64;
 Query Match
 Best Local Similarity 100.0%; Score 8; DB 1; Length 259;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1109 SLEEVQOT 1116
 DB 194 SLEEVQOT 201
 RESULT 23
 COMT RAT
 ID COMT RAT
 AC P22734;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Catechol O-methyltransferase, membrane-bound form (EC 2.1.1.6)
 DE (MB-COMT) [Contains: Catechol O-methyltransferase, soluble form
 DE (S-COMT)].
 GN COMT.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_Taxid=10116;
 RX NCBI_Taxid=10116;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94107221; PubMed=8280056;
 RT Tenhunen J., Ullmanen I.;
 "Production of rat soluble and membrane-bound catechol O-
 RT methyltransferase forms from bifunctional mRNAs";
 RL Biochem. J. 296:595-600 (1993).
 [2]
 RP SEQUENCE OF 11-264 FROM N.A.
 RX MEDLINE=91033034; PubMed=2227437;
 RA Salminen M., Lundstroem K., Tiliemann C., Savolainen R., Kalkkinen N.,
 RA Ullmanen I.;
 RT "Molecular cloning and characterization of rat liver catechol-O-
 RT methyltransferase";
 RL Gene 93:241-247 (1990).
 [3]
 RN SEQUENCE OF 1-10 FROM N.A. AND CHARACTERIZATION OF THE TWO FORMS.
 RP MEDLINE=9211472; PubMed=1765063;
 RX Ullmanen I., Lundstroem K.;

RT "Cell-free synthesis of rat and human catechol O-methyltransferase.
 RT Insertion of the membrane-bound form into microsomal membranes in
 RT vitro." Eur. J. Biochem. 202:1013-1020 (1991).
 RL
 RN
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF SOLUBLE FORM.
 RX MEDLINE=94173341; PubMed=6127373;
 RA Vldgren J., Svensson L.A., Liljas A.;
 RT "Crystal structure of catechol O-methyltransferase.";
 RL Nature 368:354-358 (1994).
 CC -1- FUNCTION: CATALYZES THE O-METHYLATION, AND THEREBY THE
 CC INACTIVATION, OF CATECHOLAMINE NEUROTRANSMITTERS AND CATECHOL
 CC HORMONES. ALSO SHORTENS THE BIOLOGICAL HALF-LIVES OF CERTAIN
 CC NEUROACTIVE DRUGS, LIKE L-DOPA, ALPHA-METHYL DOPA AND
 CC ISOPROTERENOL.
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + catechol = S-
 CC adenosyl-L-homocysteine + guaiacol.
 CC -1- COFACTOR: MAGNESIUM.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (isoform S-COMT). Type II
 CC membrane protein (isoform MB-COMT).
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a membrane-bound form/MB-COMT
 CC (shown here) and a soluble form/S-COMT; are produced by
 CC alternative initiation.
 CC -1- PTM: The N-terminus is blocked.
 CC -1- SIMILARITY: TO OTHER MAMMALIAN CATECHOL-O-METHYLTRANSFERASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z12651; CAI78276.1; -;
 DR EMBL; M60754; AAA40882.1; ALT_INIT.
 DR EMBL; M60753; AAA40881.1; ALT_INIT.
 DR PIR; JQ0787; JQ0787.
 DR PIR; S22090; S22090.
 DR PDB; 1VID; 11-JUL-96.
 DR InterPro; IPR002935; Methyltransf_3.
 DR InterPro; IPR000051; SAM_bind.
 DR Pfam; PF01596; Methyltransf_3; 1.
 KW Transferrase; Methyltransferase; Neurotransmitter degradation;
 KW Catecholamine metabolism; Transmembrane; Signal-anchor; Magnesium;
 KW Alternative initiation; 3D-structure.
 FT CHAIN 1 264 CATECHOL-O-METHYLTRANSFERASE, MEMBRANE-
 FT BOUND ISOFORM.
 FT CHAIN 45 264 CATECHOL-O-METHYLTRANSFERASE, SOLUBLE
 FT ISOFORM.
 FT INIT MET 44 44 FOR SOLUBLE ISOFORM.
 FT TRANSMEM 3 19 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT ACT_SITE 187 187
 FT ACT_SITE 242 242
 FT METAL 184 184 MAGNESIUM.
 FT METAL 212 212 MAGNESIUM.
 FT METAL 213 213 MAGNESIUM.
 SQ SEQUENCE 264 AA; 29597 MW; F535DF49C062854 CRC64;
 QY Query Match 0.7%; Score 8; DB 1; Length 264;
 DB Best Local Similarity 100.0%; Pred. No. 9.1;
 DB Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1133 GLLLLALL 1140
 DB 9 GLLLLALL 16
 RESULT 24
 ALSK_ECOLI STANDARD; PRT; 309 AA.
 ID ALSK_ECOLI
 AC P32718;

DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE D-allose kinase (EC 2.7.1.55) (Allokinase).
 GN
 OS ALSK OR B4084.
 OC Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RC MEDLINE=94089392; PubMed=8265357;
 RA Blatner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
 RA Daniels D.L.;
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
 RT region from 89.2 to 92.8 minutes.";
 RL Nucleic Acids Res. 21:5408-5417 (1993).
 RN
 RP FUNCTION.
 RC STRAIN=K12;
 RX MEDLINE=98062191; PubMed=9401019;
 RA Kim C., Song S., Park C.;
 RT "The D-allose operon of Escherichia coli K-12.";
 RL J. Bacteriol. 179:7631-7637 (1997).
 CC -1- CATALYTIC ACTIVITY: ATP + D-allose = ADP + D-allose 6-phosphate.
 CC -1- PATHWAY: D-ALLOSE METABOLISM. NOT ESSENTIAL FOR THIS PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLAR) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U00006; AAC43178.1; -;
 DR EMBL; AE000482; AAC77045.1; -;
 DR Ecocore; EGI1956; alek.
 DR InterPro; IPR000600; ROK_family.
 DR Pfam; PF00480; ROK; 1.
 DR PROSITE; PS01125; ROK; 1.
 KW Transferrase; Kinase; Carbohydrate metabolism; Complete proteome.
 KW SEQUENCE 309 AA; 33821 MW; 7C655FD2E5A8F2E CRC64;
 SQ
 QY Query Match 0.7%; Score 8; DB 1; Length 309;
 DB Best Local Similarity 100.0%; Pred. No. 10;
 DB Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 632 GAQGAAIL 639
 DB 288 GAQGAAIL 295
 RESULT 25
 NSMA_HUMAN STANDARD; PRT; 423 AA.
 ID NSMA_HUMAN
 AC O60906; O9BMR3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE sphingomyelin phosphodiesterase 2 (EC 3.1.4.12) (Neutral
 DE sphingomyelinase) (nsMase) (N-SMase) (Lyso-platelet activating factor-
 DE phospholipase C) (Lyso-PAF-PLC).
 GN SMPD2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND TISSUE SPECIFICITY.
 RX MEDLINE=98188255; PubMed=9520418;


```

RA Tomiuk S., Hofmann K., Nix M., Zumbansen M., Stoffel W.;
RT "Cloned mammalian neutral sphingomyelinase: functions in sphingolipid
RT signaling?";
RL Proc. Natl. Acad. Sci. U.S.A. 95:3638-3643 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RC Strausberg R.;
RN Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION..
RX MEDLINE=20076490; PubMed=10608884;
RA Sawai H., Dome N., Nagan N., Hannun Y.A.;
RT "Function of the cloned putative neutral sphingomyelinase as
RT lyso-platelet activating factor-phospholipase C.";
RL J. Biol. Chem. 274:38131-38139 (1999).
CC -!- FUNCTION: Converts sphingomyelin to ceramide. Hydrolyze 1-acyl-2-
CC lyso-sn-glycero-3-phosphocholine (lyso-PC) and 1-O-alkyl-2-lyso-
CC sn-glycero-3-phosphocholine (lyso-platelet activating factor). The
CC physiological substrate seems to be lyso-PAF.
CC -!- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +
CC choline phosphate.
CC -!- COFACTOR: Magnesium.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- MISCELLANEOUS: THIS PROTEIN HAS AN OPTIMUM PH OF 6.5-7.5.
CC -!- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ222801; CA10995.1; -
CC EMBL: BC000038; AA000038.1; -
CC Genew; HGNC:11121; SMPD2.
CC MIM: 603498; -
CC
CC InterPro: IPR005135; Exo_endo_phos.
CC Pfam: PF03372; Exo_endo_phos.
CC
CC Hydroxylase; Transmembrane; Magnesium.
CC TRANSMEM 330 350
CC FT TRANSMEM 334 374
CC FT METAL 49 49
CC FT SITE 180 180
CC
CC ACT_SITE 272 272
CC FT CONFLICT 3 3
CC FT SEQUENCE 423 AA; 47592 MW; 05252A923E363171 CRC64;
SQ

```

```

Query Match 0.7%; Score 8; DB 1; Length 423;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1133 GLLTLLLV 1140
DB 334 GLLTLLLV 341

```

```

RESULT 26
CPBK_MOUSE STANDARD; PRT; 470 AA.
AC 062357;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 2B20 (EC 1.14.14.1) (CYP1B20) (P24) (Fragment).
GN CYP2B20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=96428606; PubMed=8831708;
RA Damon M., Fautrel A., Marc N., Guillouzo A., Corcos L.;
RT "Isolation of a new mouse cDNA clone: hybrid form of cytochrome P450
RT 2b10 and NADPH-cytochrome P450 oxidoreductase.";
RL Biochem. Biophys. Res. Commun. 226:900-905 (1996).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES THIS ENZYME IS INVOLVED IN AN
CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound Endoplasmic reticulum.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER AS WELL AS IN KIDNEY,
CC LUNG AND INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X99715; CA68051.1; ALT_INIT.
CC DR HSSP; P00179; 1DT6.
CC DR MGD; MGI:1202389; Cyp2b20.
CC InterPro: IPR001128; Cytochrome_P450.
CC Pfam: PF00067; P450.1
CC PROSITE; PS00086; CYTOCHROME_P450.1.
CC
CC Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC Microsome; Endoplasmic reticulum; Phosphorylation.
CC FT MOD_RES 128 128
CC FT BINDING 436 436
CC FT NON_TER 470 470
CC FT SEQUENCE 470 AA; 53357 MW; 8B9CF3E2EA622642 CRC64;
SQ

```

```

Query Match 0.7%; Score 8; DB 1; Length 470;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1134 LLLTLLLV 1141
DB 6 LLLTLLLV 13

```

```

RESULT 27
CPB1_RAT STANDARD; PRT; 491 AA.
AC P00176;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 2B1 (EC 1.14.14.1) (CYP1B1) (P450-B) (P450-PB1 and
DE P450-PB2) (P450-LM2).
GN CYP2B1 OR CYP2B-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE OF 6-491 FROM N.A. (ISOZYMES PB1 AND PB2).
RX MEDLINE=82222224; PubMed=6953431;
RA Fujii-Kuriyama Y., Mizukami Y., Kawajiri K., Sogawa K., Muramatsu M.;
RT "Primary structure of a cytochrome P-450: coding nucleotide sequence
RT of phenobarbital-inducible cytochrome P-450 cDNA from rat liver.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:2793-2797 (1982).
RN [2]
RP REVISIONS TO 166; 292 AND 378 (ISOZYMES PB1 AND PB2).

```

RA Fujii-Kuriyama Y., Miukami Y., Kawajiri K., Sogawa K., Muramatsu M.,
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5443-5443(1982).
 RN [3]
 RP SEQUENCE OF 1-22.
 RA MEDLINE=7919411; Pubmed=109438;
 RX Botelho L.H., Ryan D.E., Levin W.;
 RT "Amino acid compositions and partial amino acid sequences of three
 RT highly purified forms of liver microsomal cytochrome P-450 from rats
 RT treated with polychlorinated biphenyls, phenobarbital, or 3-
 RT methylcholanthrene";
 RL J. Biol. Chem. 254:5635-5640(1979).
 RN [4]
 RP PHOSPHORYLATION.
 RX MEDLINE=90059885; Pubmed=2583091;
 RA Pyerin W., Taniguchi H.;
 RT "Phosphorylation of hepatic phenobarbital-inducible cytochrome
 RT P-450";
 RL EMBO J. 8:3003-3010(1989).
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
 CC ACIDS, AND XENOBIOTICS.
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
 CC oxidized flavoprotein + H(2)O.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
 CC -1- INDUCTION: BY PHENOBARBITAL.
 CC -1- PTM: PHOSPHORYLATION IS ACCOMPANIED BY A DECREASE IN ENZYME
 CC ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J00719; AAA41024.1; -;
 DR EMBL: M37134; AAA42028.1; -;
 DR PIR: A00176; QARTPB.
 DR HSSP: P00179; 1D76.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 KW oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum; Phosphorylation.
 FT MOD_RES 128 128 PHOSPHORYLATION (BY PKA).
 FT BINDING 436 436 HEME.
 FT VARIANT 303 303 S -> G (IN ISOZYME PB2).
 FT VARIANT 321 322 AE -> TV (IN ISOZYME PB2).
 FT VARIANT 337 337 L -> P (IN ISOZYME PB2).
 FT VARIANT 339 339 T -> S (IN ISOZYME PB2).
 FT VARIANT 344 344 S -> T (IN ISOZYME PB2).
 SQ SEQUENCE 491 AA; 55933 MW; 74615501AD5497DD CRC64;

Query Match 0.7%; Score 8; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred.No.15;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
 |||||
 Db 6 LLLALLLV 13

RESULT 28
 ID CPB2 RAT STANDARD; PRT; 491 AA.
 AC PC4157; 064582;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome P450 2B2 (EC 1.14.14.1) (CYP1B2) (P450B) (P450 PB4).
 GN CYP2B2 OR CYP2B-2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83247397; Pubmed=6306654;
 RA Mizukami Y., Sogawa K., Suwa Y., Muramatsu M., Fujii-Kuriyama Y.;
 RT "Gene structure of a phenobarbital-inducible cytochrome P-450 in rat
 RT liver";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3958-3962(1983).
 RN [2]
 RP SEQUENCE.
 RX MEDLINE=86059379; Pubmed=3877725;
 RA Frey A.B., Waxman D.J., Kreibich G.;
 RT "The structure of phenobarbital-inducible rat liver cytochrome P-450
 RT isoenzyme PB-4. Production and characterization of site-specific
 RT antibodies";
 RL J. Biol. Chem. 260:15253-15265(1985).
 RN [3]
 RP SEQUENCE OF 168-491 FROM N.A.
 RX MEDLINE=84159487; Pubmed=6689485;
 RA Phillips I.R., Shephard E.A., Ashworth A., Rabin B.R.;
 RT "Cloning and sequence analysis of a rat liver cDNA coding for a
 RT phenobarbital-inducible microheterogenous cytochrome P-450 variant:
 RT regulation of its messenger level by xenobiotics";
 RL Gene 26:41-52(1983).
 RN [4]
 RP SEQUENCE OF 281-491 FROM N.A.
 RX MEDLINE=83291091; Pubmed=6688421;
 RA Kumar A., Raphael C., Adesnik M.;
 RT "Cloned cytochrome P-450 cDNA. Nucleotide sequence and homology to
 RT multiple phenobarbital-induced mRNA species";
 RL J. Biol. Chem. 258:11280-11284(1983).
 RN [5]
 RP ERRATUM.
 RA Kumar A., Raphael C., Adesnik M.;
 RL J. Biol. Chem. 259:6039-6039(1984).
 RN [6]
 RP SEQUENCE OF 323-491 FROM N.A.
 RX MEDLINE=86205943; Pubmed=3458196;
 RA Atchison M.L., Adesnik M.;
 RT "Gene conversion in a cytochrome P-450 gene family";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:2300-2304(1986).
 RN [7]
 RP SEQUENCE OF 385-491 FROM N.A.
 RX MEDLINE=84153837; Pubmed=6322758;
 RA Affolter M., Anderson A.;
 RT "Segmental homologies in the coding and 3' non-coding sequences of
 RT rat liver cytochrome P-450e and P-450b cDNAs and cytochrome
 RT P-450e-like genes";
 RL Biochem. Biophys. Res. Commun. 118:655-662(1984).
 RN [8]
 RP SEQUENCE OF 1-20 FROM N.A.
 RX MEDLINE=88273074; Pubmed=2839467;
 RA Hashimoto T., Matsumoto T., Nishizawa M., Kawabata S.,
 RA Morohashi K., Handa S., Omura T.;
 RT "A mutant rat strain deficient in induction of a
 RT phenobarbital-inducible form of cytochrome P-450 in liver
 RT microsomes";
 RL J. Biochem. 103:487-492(1988).
 RN [9]
 RP PHOSPHORYLATION.
 RX MEDLINE=90059885; Pubmed=2583091;
 RA Pyerin W., Taniguchi H.;
 RT "Phosphorylation of hepatic phenobarbital-inducible cytochrome
 RT P-450";
 RL EMBO J. 8:3003-3010(1989).
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN

```

CC MAMPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- INDUCTION: BY PHENOBARBITAL.
CC -1- PTM: PHOSPHORYLATION IS ACCOMPANIED BY A DECREASE IN ENZYME
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, J00728; AAA41056.1; -.
CC DR EMBL, J00720; AAA41056.1; JOINED.
CC DR EMBL, J00721; AAA41056.1; JOINED.
CC DR EMBL, J00722; AAA41056.1; JOINED.
CC DR EMBL, J00723; AAA41056.1; JOINED.
CC DR EMBL, J00724; AAA41056.1; JOINED.
CC DR EMBL, J00725; AAA41056.1; JOINED.
CC DR EMBL, J00726; AAA41056.1; JOINED.
CC DR EMBL, K00996; AAA41029.1; -.
CC DR EMBL, K01626; AAA41037.1; -.
CC DR EMBL, K01721; AAA41026.1; -.
CC DR EMBL, D00250; BAA00181.1; -.
CC DR EMBL, M13234; AAA41057.1; -.
CC DR PIR, A00177; O4RP2.
CC DR PIR, A21872; A21872.
CC DR HSSP, P00179; 1DT6.
CC DR InterPro, IPR001128; Cytochrome_P450.
CC DR Pfam, PF00067; P450; 1.
CC DR PROSITE, PS00385; P450.
CC DR PROSITE, PS00086; CYTOCHROME_P450; 1.
CC KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC KM Microsome; Endoplasmic reticulum; Phosphorylation.
CC FT MOD_RES 128 128 PHOSPHORYLATION (BY PKA).
CC FT BINDING 436 436 HEME.
CC FT CONFLICT 292 292 L -> P (IN REF. 2).
CC FT CONFLICT 321 321 T -> A (IN REF. 2 AND 4).
CC FT CONFLICT 332 332 E -> V (IN REF. 1).
CC FT CONFLICT 438 438 G -> D (IN REF. 4).
CC FT CONFLICT 444 444 N -> K (IN REF. 3).
CC FT CONFLICT 473 473 K -> M (IN REF. 1).
CC FT CONFLICT 476 476 G -> D (IN REF. 2 AND 4).
CC SQ SEQUENCE 491 AA; 55932 MW; 00CB6B937FDD44BC CRC64;

Query Match 0.7%; Score 8; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
DB 6 LLLALLLV 13

RESULT 29
CPBA_MOUSE STANDARD; PRT; 500 AA.
AC P12751;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome P450 2B10 (EC 1.14.14.1) (CYP2B10) (Testosterone 16-alpha
DE hydroxylase) (P450-16-alpha) (Clone pF3/46).
GN CYP2B10 OR CYP2B-10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

CC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC NCBI_Taxid=10090;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=89118235; Pubmed=3219345;
CC RA Noshiro M., Iakso M., Kawajiri K., Negishi M.;
CC RT "Rip locus: regulation of female-specific isozyme (I-P-450(16 alpha)
CC of testosterone 16 alpha-hydroxylase in mouse liver, chromosome
CC localization, and cloning of P-450 cDNA.";
CC RL Biochemistry 27:6434-6443(1988).
CC
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
CC MAMPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
CC ACIDS, AND XENOBIOTICS.
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M21856; AAA40425.1; -.
CC DR PIR, B31047; B31047.
CC DR HSSP, P00179; 1DT6.
CC DR MGD, MGI:88598; Cyp2b10.
CC DR InterPro, IPR001128; Cytochrome_P450.
CC DR Pfam, PF00067; P450; 1.
CC DR PROSITE, PS00086; CYTOCHROME_P450; P450_NEG.
CC KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
CC KM Microsome; Endoplasmic reticulum; Phosphorylation.
CC FT MOD_RES 128 128 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
CC FT BINDING 445 445 HEME.
CC SQ SEQUENCE 500 AA; 56743 MW; P660A0DD0FBA94 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 500;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
DB 6 LLLALLLV 13

RESULT 30
ITAB_PAPCY STANDARD; PRT; 604 AA.
ID ITAB_PAPCY
AC P53711;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-1Ib (Platelet membrane glycoprotein IIb) (GpIIb IIb)
DE (GPIIb) (CD41 antigen) (Fragment).
GN ITGA2B.
OS Papio cynocephalus (yellow baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
CC NCBI_Taxid=9556;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=95129873; Pubmed=7828888;
CC RA Hayer D.J., Shoji M., Kim T.M., Runge M.S., Hanson S.R.;
CC "Alternative splicing of the mRNA encoding baboon glycoprotein
CC receptor GPIIb.";
CC RL Gene 151:267-271(1994).
CC -1- FUNCTION: INTEGRIN ALPHA-1Ib/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,

```

CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
 CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF
 CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-O-A-G-D-V IN
 CC FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
 CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
 CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
 CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL
 CC SURFACE.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
 CC DISULFIDE BOND. ALPHA-IIB ASSOCIATES WITH BETA-3.
 CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 FG-GAP REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L12233; AA65936.1; -
 CC HSSP; P06756; JUV2.
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam: PF00357; Integrin_A; 1.
 DR Pfam: PF01839; FG-GAP; 1.
 DR SMART; SM00191; Int_alpha; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 DR Platelet; Calcium.
 FT NON TER 1 1
 FT CHAIN <1 467 INTEGRIN ALPHA-IIB HEAVY CHAIN.
 FT CHAIN 468 604 INTEGRIN ALPHA-IIB LIGHT CHAIN.
 FT DOMAIN <1 558 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 559 584 POTENTIAL.
 FT DOMAIN 585 604 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 10 62 FG-GAP.
 FT CA BIND 22 30 POTENTIAL.
 FT SITE 587 591 GFFKR MOTIF.
 FT DISULFID 69 80 BY SIMILARITY.
 FT DISULFID 86 141 BY SIMILARITY.
 FT DISULFID 198 204 BY SIMILARITY.
 FT DISULFID 270 283 BY SIMILARITY.
 FT DISULFID 422 476 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 481 486 BY SIMILARITY.
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 604 AA; 66065 MW; 0B13BD1BD9E37F88 CRC64;
 Query Match 0.7%; Score 8; DB 1; Length 604;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1130 VLGGLLLL 1137
 Db 569 VLGGLLLL 576
 RESULT 31
 ZNF4 HUMAN STANDARD; PRT; 643 AA.
 ID ZNF4 HUMAN
 AC Q16587; Q9UF05; Q9UF06; Q9UF07;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein 74.
 GN ZNF4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA MEDLINE=96279060; PubMed=8663113;
 RT "The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein
 RT tightly associated with the nuclear matrix";
 RL J. Biol. Chem. 271:15458-15467(1996).
 RN [2]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA MEDLINE=94093543; PubMed=8268910;
 RA Aubry M., Demczuk S., Desmaziere C., Aikem M., Julien J.-P.,
 RA Rouleau G.A.;
 RA "Isolation of a zinc finger gene consistently deleted in DisGeorge
 RT syndrome";
 RL Hum. Mol. Genet. 2:1583-1587(1993).
 RN [3]
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RA Aubry M., Cole F.;
 RT "Alternative promoter usage and splicing of ZNF74 gene";
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RN VARIANTS LYS-117 AND 622-ASN-PHE-623.
 RX MEDLINE=21563124; PubMed=11705709;
 RA Takase K., Ohtsuki T., Migita O., Toru M., Inada T.,
 RA Yamakawa-Kobayashi K., Arinami T.;
 RT "Association of ZNF74 gene genotypes with age-at-onset of
 RT schizophrenia";
 RL Schizophrenia. Res. 52:161-165(2001).
 CC -1- FUNCTION: MAY PLAY A ROLE IN RNA METABOLISM.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2 (SHOWN HERE), 3 AND 4; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE FETAL BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X92715; CAA63379.1; -
 CC EMBL; X71623; CAA50632.1; -
 DR EMBL; AF072567; AAF21777.1; -
 DR EMBL; AF072567; AAF21777.1; JOINED.
 DR EMBL; AF072567; AAF21777.1; JOINED.
 DR EMBL; AF072567; AAF21778.1; -
 DR EMBL; AF072567; AAF21778.1; JOINED.
 DR EMBL; AF072567; AAF21778.1; JOINED.
 DR EMBL; AF072567; AAF21779.1; -
 DR EMBL; AF072567; AAF21779.1; JOINED.
 DR EMBL; AF072567; AAF21780.1; -
 DR EMBL; AF072567; AAF21780.1; JOINED.
 DR HSSP; P07248; 1PAA.
 DR GeneW; HGNC:11144; ZNF74.
 DR MIM; 194548; -
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam; PF00096; Zf-C2H2; 12.
 DR Pfam; PF01352; KRAB; 1.
 DR PRINTS; PR00048; ZINCFINGER.
 DR ProDom; PD000003; Znf_C2H2; 9.
 DR SMART; SM00349; KRAB; 1.
 DR SMART; SM00355; Znf_C2H2; 12.
 DR PROSITE; PS50805; KRAB; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.

RA	Lanxin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Lueros U.S., Maitri R., Marziani A., Militscher U., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.B., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utterback T., Van Aken S., Vaysberg M., Vysockakova V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter U.C., Davis R.W.; "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
RT	Nature 408:816-820(2000).
RL	- FUNCTION: PROBABLE RECEPTOR.
CC	- SUBCELLULAR LOCATION: Type I membrane protein.
CC	- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC	- PMB: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.
CC	- SIMILARITY: BELONGS TO THE SSR/TRR FAMILY OF PROTEIN KINASES.
CC	- SIMILARITY: CONTAINS 9 LEUCINE-RICH REPEATS (LRR).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outpost - the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). -----
CC	EMBL; L00670; AAA32876.1; -.
DR	EMBL; AC026480; AACG1302.1; -.
DR	HSSP; PI2931; IFMK.
DR	InferPro; IPFR00719; Euk_Pkinase.
DR	InferPro; IPRO01611; LKR.
DR	InferPro; IPRO03592; LRR_out.
DR	InferPro; IPRO03591; LRR_Typ.
DR	InferPro; IPRO04040; STY_Pkinase.
DR	InferPro; IPRO02290; Ser_Chk_Pkinase.
DR	Pfam; PF00063; pkinase; 1.
DR	Pfam; Pf00560; LRR_11.
DR	PIRDom; PD000001; Euk_pkinase; 1.
DR	SMART; SMART0370; LRR_7.
DR	SMART; SMART0369; LRR_TYP; 1.
DR	SMART; SMART0221; STYC; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
KW	Transferrase; Serine/threonine-protein kinase; ATP-binding; Transmembrane Receptor; Glycoprotein; Signal; Repeat; Leucine-rich repeat; Phosphorylation.
KM	Leucine-rich repeat; phosphorylation.
FT	SIGNAL 1..23 POTENTIAL.
FT	CHAIN 24..942 PUTATIVE RECEPTOR PROTEIN KINASE TMK1.
FT	DOMAIN 24..482 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 483..503 POTENTIAL.
FT	DOMAIN 504..942 CYTOPLASMIC (POTENTIAL).
FT	REPEAT 87..110 LRR 1.
FT	REPEAT 111..133 LRR 2.
FT	REPEAT 135..159 LRR 3.
FT	REPEAT 185..209 LRR 4.
FT	REPEAT 231..253 LRR 5.
FT	REPEAT 254..278 LRR 6.
FT	REPEAT 280..300 LRR 7.
FT	REPEAT 385..408 LRR 8.
FT	REPEAT 409..436 LRR 9.
FT	DOMAIN 588..869 PROTEIN KINASE.
FT	NP_BIND 594..602 ATP (BY SIMILARITY).
FT	BINDING 616..616 ATP (BY SIMILARITY).
FT	ACT_SITE 717..717 BY SIMILARITY.
FT	CARBOHYD 86..86 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 99..99 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 158..158 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 164..164 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 171..171 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 230..230 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 285..285 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 363..363 N-LINKED (GLCNAC...) (POTENTIAL).

```

FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 942 AA; 102387 MW; 938300B52FF549DE CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1128 GSVLGGL 1135
   |||||
Db 486 GSVLGGL 493

RESULT 33
VP2_BTV3V STANDARD; PRT; 959 AA.
AC 006598;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer capsid protein VP2.
GN S2.
OS Bluetongue virus (serotype 3 / isolate South Africa-vaccine).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=36424;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021485; PubMed=2171239;
RA Gould A.R., Pritchard L.I.;
RT "Relationships amongst bluetongue viruses revealed by comparisons of
RT capsid and outer coat protein nucleotide sequences.";
RL Virus Res. 17:31-52(1990).
CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
CC MAJOR CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -1- SIMILARITY: BELONGS TO THE REOVIRUSES VP2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X55801; CAA39323.1; -
CC PIR; B60017; B60017.
DR InterPro; IPR001742; Orb1_VP2.
DR Pfam; PF00898; Orb1_VP2; 1.
DR ProDom; PD002938; Orb1_VP2; 1.
KW Coat protein.
SQ SEQUENCE 959 AA; 112163 MW; 41B230E31803588F CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1159 EKREKLE 1166
   |||||
Db 798 EKREKLE 805

RESULT 34
ITAB MOUSE STANDARD; PRT; 1033 AA.
AC 090200; Q9Z2M0; Q64228;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-IIB precursor (Platelet membrane glycoprotein IIB)
DE (GpIb alpha IIB) (GPIIb) (CD41 antigen).
GN ITGA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20042231; PubMed=10572112;
RA Thornton M.A., Poncz M.;
RT "Characterization of the murine platelet alphaIIB gene and encoded
RT cDNA.";
RL Blood 94:3947-3950(1999).
RN [2]
RP SEQUENCE OF 805-865 FROM N.A.
RX MEDLINE=92381049; PubMed=1512266;
RA Chen Y.Q., Gao X., Tamar J., Tang D., Grossi I.M., Chelladurai M.,
RA Kunicki T.J., Fligiel S.E., Taylor J.D., Honn K.V.;
RT "Identification of the alpha IIB beta 3 integrin in murine tumor
RT cells.";
RL J. Biol. Chem. 267:17314-17320(1992).
RN [3]
RP SEQUENCE OF 790-1022 FROM N.A.
RA Rout U.K., Arment D.R.;
RT Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-484 FROM N.A.
RX MEDLINE=20179880; PubMed=10713093;
RA Puzon-McLaughlin W., Kamata T., Takada Y.;
RT "Multiple discontinuous ligand-mimetic antibody binding sites define a
RT ligand binding pocket in integrin alphaIIb beta3.";
RL J. Biol. Chem. 275:7795-7802(2000).
CC -1- FUNCTION: INTEGRIN ALPHA-IIB/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,
CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF
CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-Q-A-G-D-V IN
CC FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL
CC SURFACE.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-IIB ASSOCIATES WITH BETA-3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF169829; AAF06996.1; -
CC EMBL; AF170316; AAD56216.1; -
CC EMBL; S43388; AAB23054.2; -
CC EMBL; AF045019; AAB20339.1; -
CC EMBL; AF166384; AAF43997.1; -
CC HSSP; P06756; 1JY2.
DR MGD; MGI:96601; Itga2b.
DR InterPro; IPR00413; Integrin_alpha.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF01839; FG-GAP; 5.
DR PRINTS; P01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Repeat; Calcium.
FT SIGNAL 1 31 BY SIMILARITY.
FT CHAIN 32 1033 INTEGRIN ALPHA-IIB.
FT CHAIN 32 1033 INTEGRIN ALPHA-IIB HEAVY CHAIN (BY
FT CHAIN 32 1033 SIMILARITY).
FT CHAIN ? 1033 INTEGRIN ALPHA-IIB LIGHT CHAIN (BY
FT SIMILARITY).

```

```

FT DOMAIN 32 988 EXTRACELLULAR (POTENTIAL).
FT TRAMSEM 989 1014 POTENTIAL.
FT DOMAIN 1015 1033 CYTOPLASMIC (POTENTIAL).
FT REPEAT 47 107 FG-GAP 1.
FT REPEAT ? ? FG-GAP 2.
FT REPEAT ? ? FG-GAP 3.
FT REPEAT 262 315 FG-GAP 4.
FT REPEAT 316 383 FG-GAP 5.
FT REPEAT 384 443 FG-GAP 6.
FT REPEAT 444 496 FG-GAP 7.
FT CA_BIND 273 328 POTENTIAL.
FT CA_BIND 327 335 POTENTIAL.
FT CA_BIND 395 403 POTENTIAL.
FT CA_BIND 456 464 POTENTIAL.
FT SITE 1017 1021 GEFKR MOTIF.
FT DISULFID 87 96 BY SIMILARITY.
FT DISULFID 138 161 BY SIMILARITY.
FT DISULFID 177 197 BY SIMILARITY.
FT DISULFID 503 514 BY SIMILARITY.
FT DISULFID 520 575 BY SIMILARITY.
FT DISULFID 632 638 BY SIMILARITY.
FT DISULFID 704 717 BY SIMILARITY.
FT DISULFID 856 905 INTERCHAIN (BY SIMILARITY).
FT DISULFID 911 916 BY SIMILARITY.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 710 710 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 957 957 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 236 236 T -> S (IN REF. 4).
FT CONFLICT 270 270 S -> A (IN REF. 4).
FT CONFLICT 464 464 D -> G (IN REF. 4).
FT CONFLICT 471 471 W -> G (IN REF. 4).
FT CONFLICT 483 483 G -> V (IN REF. 4).
FT CONFLICT 805 805 S -> R (IN REF. 2 AND 3).
FT CONFLICT 834 834 P -> H (IN REF. 3).
FT CONFLICT 848 848 VQ -> LR (IN REF. 3).
FT CONFLICT 865 865 D -> E (IN REF. 2).
FT CONFLICT 933 933 A -> V (IN REF. 3).
SQ SEQUENCE 1033 AA; 112697 MW; 7F20A16C8EF2BD7 CRC64;

```

```

Query Match 0.7%; Score 8; DB 1; Length 1033;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1130 VVGGLLL 1137
Db 999 VVGGLLL 1006

```

```

RESULT 35
ITAB_HUMAN STANDARD; PRT; 1039 AA.
AC P08514; Q14443; G95366;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integrin alpha-IIb precursor (Platelet membrane glycoprotein IIb)
DE (GPIIb) (GPIIb) (CD41 antigen).
GN ITGA2B OR ITGAB OR GP2B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=87250457; PubMed=2439501;
RA Poncz M., Eisman R., Heideneich R., Silver S.M., Vilaire G.,
RA Surety S., Schwartz E., Bennett J.S.;
RT "Structure of the platelet membrane glycoprotein IIb. Homology to the
RT alpha subunits of the vitronectin and fibronectin membrane
RT receptors."
RL J. Biol. Chem. 262:8476-8482(1987).

```

```

RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=90265363; PubMed=2345548;
RA Frachet P., Uzan G., Thevenon D., Denarier E., Prandini M.H.,
RA Marguerie G.;
RT "GPIIb and GPIIa amino acid sequences deduced from human
RT megakaryocyte cDNAs."
RL Mol. Biol. Rep. 14:27-33(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=90212612; PubMed=2322558;
RA Heideneich R., Eisman R., Surety S., Delgrosso K., Bennett J.S.,
RA Schwartz E., Poncz M.;
RT "Organization of the gene for platelet glycoprotein IIb."
RL Biochemistry 29:1232-1244(1990).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Erythrocyte;
RX MEDLINE=90277633; PubMed=2351656;
RA Bray P.F., Leung C.S.-I., Shuman M.A.;
RT "Human platelets and megakaryocytes contain alternately spliced
RT glycoprotein IIb mRNAs."
RL U. Biol. Chem. 265:9587-9590(1990).
RN [5]
RP SEQUENCE OF 32-56 AND 903-917.
RX MEDLINE=87041455; PubMed=3534886;
RA Charo I.F., Fitzgerald L.A., Steiner B., Rall S.C., Bekeart L.S.,
RA Phillips D.R.;
RT "Platelet glycoproteins IIb and IIla: evidence for a family of
RT immunologically and structurally related glycoproteins in mammalian
RT cells."
RL Proc. Natl. Acad. Sci. U.S.A. 83:8351-8355(1986).
RN [6]
RP SEQUENCE OF 392-1039 FROM N.A. (ISOFORM 1).
RX MEDLINE=8811709; PubMed=3422188;
RA Uzan G., Frachet P., Lajmanovich A., Prandini M.H., Denarier E.,
RA Duperray A., Loftus J., Ginsberg M., Plow E., Marguerie G.;
RT "cDNA clones for human platelet GPIIb corresponding to mRNA from
RT megakaryocytes and HBL cells. Evidence for an extensive homology to
RT other Arg-Gly-Asp adhesion receptors."
RL Eur. J. Biochem. 171:87-93(1988).
RN [7]
RP SEQUENCE OF 868-1039 FROM N.A. (ISOFORM 1).
RX MEDLINE=88059639; PubMed=3479442;
RA Bray P.F., Rosa J.P., Johnston G.I., Shiu D.T., Cook R.G., Lau C.,
RA Kan Y.W., McEvers R.P., Shuman M.A.;
RT "Platelet glycoprotein IIb. Chromosomal localization and tissue
RT expression."
RL J. Clin. Invest. 80:1812-1817(1987).
RN [8]
RP SEQUENCE OF 1-62 AND 1021-1039 FROM N.A.
RX MEDLINE=89025907; PubMed=2845986;
RA Prandini M.H., Denarier E., Frachet P., Uzan G., Marguerie G.;
RT "Isolation of the human platelet glycoprotein IIb gene and
RT characterization of the 5' flanking region."
RL Biochem. Biophys. Res. Commun. 156:595-601(1988).
RN [9]
RP SEQUENCE OF 487-501 AND 1026-1038.
RX MEDLINE=87101510; PubMed=3801670;
RA Hiraiwa A., Matsukage A., Shiku H., Takahashi T., Naito K., Yamada K.;
RT "Purification and partial amino acid sequence of human platelet
RT membrane glycoproteins IIb and IIla."
RL Blood 69:560-564(1987).
RN [10]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=89374157; PubMed=2775232;
RA Calvete J.J., Henschen A., Gonzalez-Rodriguez J.;
RT "Complete localization of the intrachain disulphide bonds and the N-
RT glycosylation points in the alpha subunit of human platelet
RT glycoprotein IIb."
RL Biochem. J. 261:561-568(1989).
RN [11]
RP PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITE SER-874.

```

RX MEDLINE=93345693; PubMed=7686323;
 RA Calvete J.U., Muniz-Diaz E.;
 RT "Localization of an O-glycosylation site in the alpha-subunit of the
 human platelet integrin GPIIb/IIIa involved in Baka (HPA-3a)
 RT allonitigen expression.";
 RL FEBS Lett. 328:30-34(1993).
 RN [12]
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 3).
 RX MEDLINE=9025846; PubMed=9809974;
 RA Trikha M., Cai Y., Grignon D., Honn K.V.;
 RT "Identification of a novel truncated alphaIIb integrin.";
 RL Cancer Res. 58:4771-4775(1998).
 RN [13]
 RP MUTAGENESIS OF PRO-1029 AND PRO-1030.
 RX MEDLINE=99230329; PubMed=10212286;
 RA Leisner T.M., Wencel-Drake J.D., Wang W., Lam S.C.;
 RT "Bidirectional transmembrane modulation of integrin alphaIIb beta3
 RT conformational.";
 RL J. Biol. Chem. 274:12945-12949(1999).
 RN [14]
 RP VARIANT HPA-3 (BAK).
 RX MEDLINE=90275262; PubMed=2350579;
 RA Lyman S., Acker R.H., Visentin G.P., Newman P.J.;
 RT "Polymorphism of human platelet membrane glycoprotein IIb associated
 RL Blood 75:2343-2348(1990).
 RN [15]
 RP VARIANT GTA ASP-273.
 RX MEDLINE=94110321; PubMed=8282784;
 RA Poncz M., Rifat S., Collier B.S., Newman P.J., Shattil S.J.,
 RA Parrella T., Fortina P., Bennett J.S.;
 RT "Glanzmann thrombasthenia secondary to a Gly273-->ASP mutation
 RT adjacent to the first calcium-binding domain of platelet glycoprotein
 RT IIb.";
 RL J. Clin. Invest. 93:172-179(1994).
 RN [16]
 RP VARIANT GTA ASP-449.
 RX MEDLINE=94140878; PubMed=7508443;
 RA Wilcox D.A., Mautler J.-L., Pidard D., Newman P.J.;
 RT "A single amino acid substitution flanking the fourth calcium binding
 RT domain of alpha IIb prevents maturation of the alpha IIb beta 3
 RT integrin complex.";
 RL J. Biol. Chem. 269:4450-4457(1994).
 RN [17]
 RP VARIANT GTA HIS-358.
 RX MEDLINE=95221604; PubMed=7706461;
 RA Wilcox D.A., Paddock C.M., Lyman S., Gill J.C., Newman P.J.;
 RT "Glanzmann thrombasthenia resulting from a single amino acid
 RT substitution between the second and third calcium-binding domains of
 RT GPIIb. Role of the GPIIb amino terminus in integrin subunit
 RT association.";
 RL J. Clin. Invest. 95:1553-1560(1995).
 RN [18]
 RP REVIEW ON GTA VARIANTS.
 RX MEDLINE=95184171; PubMed=7878622;
 RA Bray P.F.;
 RT "Inherited diseases of platelet glycoproteins: considerations for
 RT rapid molecular characterization.";
 RL Thromb. Haemost. 72:492-502(1994).
 RN [19]
 RP VARIANT GTA PRO-778.
 RX MEDLINE=98438330; PubMed=9763559;
 RA Tadokoro S., Tomiyama Y., Honda S., Arai M., Yamamoto N., Shitaga M.,
 RA Kosugi S., Kanakura Y., Kurata Y., Matsuzawa Y.;
 RT "A Glu747-->Pro substitution in the IIb subunit is responsible for a
 RT moderate IIb beta3 deficiency in Glanzmann thrombasthenia.";
 RL Blood 92:2750-2758(1998).
 RN [20]
 RP VARIANTS GTA SER-320; LYS-355 AND PRO-778.
 RX MEDLINE=98387769; PubMed=9722314;
 RA Ambo H., Kanata T., Honda M., Kawai Y., Oda A., Murata M., Takada Y.,
 RA Ikeda Y.;
 RT "Novel point mutations in the alphaIIb subunit (Phe289-->Ser,

RT Glu324-->Lys and Glu747-->Pro) causing thrombasthenic phenotypes in
 RT four Japanese patients.";
 RL Br. J. Haematol. 102:829-840(1998).
 RN [21]
 RP VARIANTS GTA ALA-176 AND LEU-176.
 RX MEDLINE=20076256; PubMed=10607701;
 RA Basant R.B., French D.L., Vialtre G., Brown D.L., Chen F.,
 RA Collier B.S., Derick J.M., Garner T.K., Bennett J.S., Poncz M.;
 RT "A naturally occurring mutation near the amino terminus of alphaIIb
 RT defines a new region involved in ligand binding to alphaIIb beta3.";
 RL Blood 95:180-188(2000).
 RN [22]
 RP FUNCTION: INTEGRIN ALPHA-IIIB/BETA-3 IS A RECEPTOR FOR FIBRONECTIN,
 CC FIBRINOGEN, PLASMINOGEN, PROTHROMBIN, THROMBOSPONDIN AND
 CC VITRONECTIN. IT RECOGNIZES THE SEQUENCE R-G-D IN A WIDE ARRAY OF
 CC LIGANDS. IT RECOGNIZES THE SEQUENCE H-H-L-G-G-A-K-Q-A-G-D-V IN
 CC FIBRINOGEN GAMMA CHAIN. FOLLOWING ACTIVATION INTEGRIN ALPHA-
 CC IIB/BETA-3 BRINGS ABOUT PLATELET/PLATELET INTERACTION THROUGH
 CC BINDING OF SOLUBLE FIBRINOGEN. THIS STEP LEADS TO RAPID PLATELET
 CC AGGREGATION WHICH PHYSICALLY PLUGS RUPTURED ENDOTHELIAL CELL
 CC SURFACE.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
 CC DISULFIDE BOND. ALPHA-IIIB ASSOCIATES WITH BETA-3.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: ISOFORM 1 AND ISOFORM 2 WERE IDENTIFIED IN
 CC PLATELETS AND MEGACARYOCYTES, BUT NOT IN RETICULOCYTES OR IN
 CC JURKAT AND U937 WHITE BLOOD CELL LINE. ISOFORM 3 IS EXPRESSED BY
 CC LEUCEMIA, PROSTATE ADENOCARCINOMA AND MELANOMA CELLS BUT NOT BY
 CC PLATELETS OR NORMAL PROSTATE OR BREAST EPITHELIAL CELLS.
 CC -1- POLYMORPHISM: POSITION 874 IS ASSOCIATED WITH PLATELET-SPECIFIC
 CC ALLOANTIGEN HPA-3/BAK/LEK. HPA-3/BAK(A)/LEK(A) HAS ILE-874 AND

Query Match 0.7%; Score 8; DB 1; Length 1039;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1130 VLGGLLL 1137
 Db 1004 VLGGLLL 1011

RESULT 36
 ITAS XENLA STANDARD; PRT: 1050 AA.
 AC 006274;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-5 precursor (Fibronectin receptor alpha subunit)
 DE (Integrin alpha-P) (VLA-5).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95344994; PubMed=7619730;
 RA Joce T.O., Whitaker C.A., Meng F., Desimone D.W., Grau V.,
 RA Hausen P.;
 RT "Integrin alpha 5 during early development of Xenopus laevis.";
 RL Mech. Dev. 50:187-199(1995).
 RN [2]
 RP SEQUENCE OF 318-393 FROM N.A.
 RX MEDLINE=94008528; PubMed=8404528;
 RA Whitaker C.A., Desimone D.W.;
 RT "Integrin alpha subunit mRNAs are differentially expressed in early
 RT Xenopus embryos.";
 RL Development 117:1239-1249(1993).
 CC -1- FUNCTION: INTEGRIN ALPHA-5/BETA-1 IS A RECEPTOR FOR FIBRONECTIN.
 CC IT RECOGNIZES THE SEQUENCE R-G-D IN ITS LIGANDS.

CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
 CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
 CC DISULFIDE BOND. ALPHA-5 ASSOCIATES WITH BETA-1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).
 CC -----
 CC EMBL: U12683; AAA99668.1; -
 CC HSP: P06756; IUV2.
 CC InterPro: IPR000413; Integrin_alpha.
 CC Pfam: PF00357; Integrin_A; 1.
 CC Pfam: PF01839; FG-GAP; 5.
 CC PRINTS: PR01185; INTEGRINA.
 CC SMART: SM00191; Int_alpha; 5.
 CC PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 CC Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 CC Signal; Repeat.
 CC SIGNAL: 1
 CC CHAIN: 33 1050 POTENTIAL. INTEGRIN ALPHA-5.
 CC CHAIN: 33 932 INTEGRIN ALPHA-5 HEAVY CHAIN (POTENTIAL).
 CC CHAIN: 933 1050 INTEGRIN ALPHA-5 LIGHT CHAIN (POTENTIAL).
 CC DOMAIN: 33 997 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM: 997 1022 POTENTIAL.
 CC DOMAIN: 1023 1050 CYTOPLASMIC (POTENTIAL).
 CC REPEAT: 48 110 FG-GAP 1.
 CC REPEAT: 120 188 FG-GAP 2.
 CC REPEAT: 189 246 FG-GAP 3.
 CC REPEAT: 259 312 FG-GAP 4.
 CC REPEAT: 313 373 FG-GAP 5.
 CC REPEAT: 379 438 FG-GAP 6.
 CC REPEAT: 442 494 FG-GAP 7.
 CC REPEAT: 324 332 POTENTIAL.
 CC CA_BIND: 390 398 POTENTIAL.
 CC CA_BIND: 454 462 POTENTIAL.
 CC SITE: 1025 1029 GEFKR MOTIF.
 CC DISULFID: 90 99 BY SIMILARITY.
 CC DISULFID: 145 166 BY SIMILARITY.
 CC DISULFID: 182 195 BY SIMILARITY.
 CC DISULFID: 502 513 BY SIMILARITY.
 CC DISULFID: 519 575 BY SIMILARITY.
 CC DISULFID: 636 642 BY SIMILARITY.
 CC DISULFID: 708 721 BY SIMILARITY.
 CC DISULFID: 862 910 INTERCHAIN (BY SIMILARITY).
 CC DISULFID: 912 922 BY SIMILARITY.
 CC CARBOHYD: 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 763 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD: 861 861 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE: 1050 AA; 115961 MW; 10ED9615358BD918 CRC64;

Query Match 0.7%; Score 8; DB 1; Length 1050;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GILLTALL 1140
 DB 1010 GILLTALL 1017

RESULT 37
 ID ITAI DROME STANDARD; PRT: 1146 AA.
 AC 02427; 09YV6;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Integrin alpha-PS1 precursor (position-specific antigen 1, alpha
 DE chain) (Protein multiple edematous wings).
 GN MEW OR CG1771.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Oregon-R;
 RX MEDLINE=94059764; PubMed=8240969;
 RT Wehrli M., Diantonio A., Fearnley I.M., Smith R.J., Wilcox M.;
 "Cloning and characterization of alpha PS1, a novel Drosophila
 RT melanogaster integrin.";
 RL Mech. Dev. 43:121-36(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Suton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chang M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helz G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhattacharya D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Ewanigista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalaali M., Kalush F., Karpen G.H., Ke Z., Kennison U.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Part V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtkars R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.C., Wu D., Yang S., Yao Q.A.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- FUNCTION: INTEGRIN ALPHA-PS1/BETA-PS IS A RECEPTOR FOR LAMININ.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RX SEQUENCE FROM N.A.
 RA MEDLINE=8166645; PubMed=3327687;
 RA Corbi A.L., Miller L.J., O'Connor K., Larson R.S., Springer T.A.;
 RT "cDNA cloning and complete primary structure of the alpha subunit of
 a leukocyte adhesion glycoprotein, p150,95.";
 RL EMBL J. 6:4023-4028(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90153906; PubMed=2303426;
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
 RT "Genomic structure of an integrin alpha subunit, the leukocyte
 p150,95 molecule.";
 RL J. Biol. Chem. 265:2782-2788(1990).
 RN [3]
 RP ERRATUM.
 RA Corbi A.L., Garcia-Aguilar J., Springer T.A.;
 RL J. Biol. Chem. 265:12750-12751(1990).
 RN [4]
 RP SEQUENCE OF 20-43.
 RX MEDLINE=87167596; PubMed=3549901;
 RX Miller L.J., Wiebe M., Springer T.A.;
 RT "Purification and alpha subunit N-terminal sequences of human Mac-1
 and p150,95 leukocyte adhesion proteins.";
 RL J. Immunol. 138:2381-2383(1987).
 CC -1- FUNCTION: INTEGRIN ALPHA-X/BETA-2 IS A RECEPTOR FOR FIBRINOGEN. IT
 CC RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN. IT MEDIATES CELL-CELL
 CC INTERACTION DURING INFLAMMATORY RESPONSES. IT IS ESPECIALLY
 CC IMPORTANT IN MONOCYTE ADHESION AND CHEMOTAXIS.
 CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. ALPHA-X
 CC ASSOCIATES WITH BETA-2.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN MONOCYTES AND
 CC GRANULOCYTES.
 CC -1- DOMAIN: THE INTEGRIN I-DOMAIN (INSERT) IS A VFMA DOMAIN. INTEGRINS
 CC WITH I-DOMAINS DO NOT UNDERGO PROTEASE CLEAVAGE.
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VFMA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 7 FG-GAP REPEATS.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD11c entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd11c.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M81695; AAA59180.1; -;
 DR EMBL; Y00093; CAA68283.1; -;
 DR EMBL; M29165; -; NOT ANNOTATED CDS.
 DR EMBL; M29487; AAA51620.1; ALT SEQ.
 DR EMBL; M29482; AAA51620.1; JOINED.
 DR EMBL; M29483; AAA51620.1; JOINED.
 DR EMBL; M29484; AAA51620.1; JOINED.
 DR EMBL; M29485; AAA51620.1; JOINED.
 DR EMBL; M29486; AAA51620.1; JOINED.
 DR PIR; A35584; RWHUIC.
 DR HSSP; P11215; 1A8X.
 DR Genew; HGNC; 6152; ITGAX.
 DR MIM; 151510; -;
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF00037; vwa; 1.
 DR Pfam; PF01839; FG-GAP_5.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.

DR SMART; SM00191; Int_alpha; 4.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VFMA; 1.
 KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
 KW Signal; Magnesium; Calcium; Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 1163
 FT DOMAIN 20 1107
 FT TRANSMEM 1108 1128
 FT DOMAIN 1129 1163
 FT REPEAT 34 87
 FT REPEAT 34 87
 FT REPEAT 165 351
 FT DOMAIN 165 351
 FT REPEAT 402 453
 FT REPEAT 455 517
 FT REPEAT 518 576
 FT REPEAT 581 633
 FT CA_BIND 466 474
 FT CA_BIND 530 538
 FT CA_BIND 593 601
 FT SITE 1131 1135
 FT DISULFID 69 76
 FT DISULFID 108 126
 FT DISULFID 655 712
 FT DISULFID 771 777
 FT DISULFID 848 863
 FT DISULFID 998 1022
 FT DISULFID 1027 1032
 FT CARBOHYD 61 61
 FT CARBOHYD 89 89
 FT CARBOHYD 332 392
 FT CARBOHYD 697 697
 FT CARBOHYD 735 735
 FT CARBOHYD 839 839
 FT CARBOHYD 939 939
 FT CARBOHYD 1050 1050
 FT CONFLICT 490 490
 FT CONFLICT 756 756
 SQ SEQUENCE 1163 AA; 12788 MW; 6C4E19CC3F82A473 CRC64;
 Query Match 0.7%; Score 8; DB 1; Length 1163;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1132 GGLLLAL 1139
 DB 1115 GGLLLAL 1122
 RESULT 40
 CA36 HUMAN
 ID CA36_HUMAN STANDARD; PRT; 3176 AA.
 AC P12111; Q16501;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Collagen alpha 3(VI) chain precursor.
 GN COL6A3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=fibroblast;
 RX MEDLINE=90151612; PubMed=1689238;
 RX Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,
 RA Glaville R., Mayer U., Mann K., Deutzmann R., Timpl R.;
 RT "Mosaic structure of globular domains in the human type VI collagen
 alpha 3 chain: similarity to von Willebrand factor, fibronectin,
 RT actin, salivary proteins and apolipoprotein type protease inhibitors.";

RL EMBL J. 9:385-393(1990).
RN [2]
RN REVIEWS.
RA Chu M.-L.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2038-2373 FROM N.A.
RX MEDLINE=89066644; PubMed=3198591;
RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,
RT Timpl R.;
RT "Amino acid sequence of the triple-helical domain of human collagen
RT type VI";
RL J. Biol. Chem. 263:18601-18606(1988).
RN [4]
RP SEQUENCE OF 2092-2157 FROM N.A.
RX MEDLINE=88029444; PubMed=3665927;
RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,
RA Hsu-Chen C.-C., Bernard M.P., Timpl R.;
RT "Characterization of three constituent chains of collagen type VI by
RT peptide sequences and cDNA clones";
RL Eur. J. Biochem. 168:309-317(1987).
RN [5]
RP SEQUENCE OF 2092-2151 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88161046; PubMed=3348212;
RA Weil D., Mattei M.-G., Passagge E., van Cong N., Pribula-Conway D.,
RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal localization of human genes encoding the
RT three chains of type VI collagen";
RL Am. J. Hum. Genet. 42:435-445(1988).
RN [6]
RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=93054780; PubMed=1339440;
RA Zanussi S., Doliana R., Segat D., Bonaldo P., Colombatti A.;
RT "The human type VI collagen gene. mRNA and protein variants of the
RT alpha 3 chain generated by alternative splicing of an additional 5-end
RT exon";
RL J. Biol. Chem. 267:24082-24089(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.
RX MEDLINE=95182468; PubMed=7533217;
RA Arnoux B., Merigieu K., Saludjian P., Norris F., Norris K., Bjoern S.,
RA Olsen O., Petersen L., Ducruix A.;
RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of
RT human type VI collagen";
RL J. Mol. Biol. 246:609-617(1995).
RN [8]
RP STRUCTURE BY NMR OF 3102-3164.
RX MEDLINE=96398604; PubMed=8805527;
RA Zweckstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,
RA Holak T.A.;
RT "Structure and multiple conformations of the kunitz-type domain from
RT human type VI collagen alpha3(VI) chain in solution";
RL Structure 4:195-209(1996).
RN [9]
RP STRUCTURE BY NMR OF 3107-3164.
RX MEDLINE=97410331; PubMed=9265624;
RA Soerensen M.D., Bjorn S., Norris K., Olsen O., Petersen L.,
RA James T.L., Led J.J.;
RT "Solution structure and backbone dynamics of the human alpha3-chain
RT type VI collagen C-terminal Kunitz domain";
RL Biochemistry 36:10439-10450(1997).
RN [10]
RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.
RX MEDLINE=98204804; PubMed=9536084;
RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,
RA Stajich J.M., Viles K., Vance J.M., Chu M.-L., Speer M.C.;
RT "Missense mutation in a von Willebrand factor type A domain of the
RT alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem
RT myopathy";
RL Hum. Mol. Genet. 7:807-812(1998).
CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),

CC ALPHA 2(VI), AND ALPHA 3(VI).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- DISEASE: DEFECTS IN COL6A3 ARE A CAUSE OF BETHLEM MYOPATHY (BM).
CC BM IS A RARE AUTOSOMAL DOMINANT PROXIMAL MYOPATHY CHARACTERIZED BY
CC EARLY CHILDHOOD ONSET (COMPLETE PENETRANCE BY THE AGE OF 5) AND
CC JOINT CONTRACTURES MOST FREQUENTLY AFFECTING THE ELBOWS AND
CC ANKLES.
CC -!- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 12 WFPA DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X52022; CAA36267.1; -;
CC EMBL; X06196; CAA29557.1; -;
CC EMBL; M20778; -; NOT ANNOTATED_CDS.
CC EMBL; M27449; AAA52057.1; -;
CC EMBL; S49432; AAB24261.1; -;
CC PIR; C31952; C31952.
CC PDB; 1KNT; 01-NOV-94.
CC PDB; 2KNT; 15-MAY-97.
CC PDB; 1KUN; 12-NOV-97.
CC Genew; HGNC:2213; COL6A3.
CC MIM; 120250; -;
CC MIM; 158810; -;
CC InterPro; IPR000087; Collagen.
CC InterPro; IPR003961; FN_III.
CC InterPro; IPR002223; Kunitz_BPTI.
CC InterPro; IPR002035; WFP A.
CC Pfam; PF00014; Kunitz_BPTI; 1.
CC Pfam; PF00092; wfa; 11.
CC Pfam; PF01391; Collagen; 5.
CC PRINTS; PR00759; BASICPTASE.
CC PRODOM; PD000222; Kunitz_BPTI; 1.
CC SMART; SM00060; FN3; 1.
CC SMART; SM00131; KU; 1.
CC SMART; SM00327; WFA; 12.
CC PROSITE; PS00280; BPTI_KUNITZ_1; 1.
CC PROSITE; PS00279; BPTI_KUNITZ_2; 1.
CC PROSITE; PS50234; WFP A; 12.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;
KW Signal; 3d-structure; Disease mutation; Polymorphism;
KW Alternative splicing.
CC FT SIGNAL 1 25 POTENTIAL.
CC FT CHAIN 26 3176 COLLAGEN ALPHA 3(VI) CHAIN.
CC FT DOMAIN 26 2038 NONHELICAL REGION.
CC FT DOMAIN 2039 2375 TRIPLE-HELICAL REGION.
CC FT DOMAIN 2376 3176 NONHELICAL REGION.
CC FT DOMAIN 39 213 WFP A 1.
CC FT DOMAIN 242 419 WFP A 2.
CC FT DOMAIN 445 620 WFP A 3.
CC FT DOMAIN 639 816 WFP A 4.
CC FT DOMAIN 837 1009 WFP A 5.
CC FT DOMAIN 1029 1205 WFP A 6.
CC FT DOMAIN 1233 1404 WFP A 7.
CC FT DOMAIN 1436 1609 WFP A 8.
CC FT DOMAIN 1639 1812 WFP A 9.
CC FT DOMAIN 1838 2024 WFP A 10.
CC FT DOMAIN 2402 2581 WFP A 11.
CC FT DOMAIN 2619 2815 WFP A 12.
CC FT DOMAIN 2967 3076 FIBRONECTIN TYPE-III.
CC FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.

```

FT SITE 2040 2042 CELL ATTACHMENT SITE.
FT SITE 2136 2138 CELL ATTACHMENT SITE.
FT SITE 2148 2150 CELL ATTACHMENT SITE.
FT SITE 2154 2156 CELL ATTACHMENT SITE.
FT SITE 2370 2372 CELL ATTACHMENT SITE.
FT ACT_SITE 3121 3122 REACTIVE BOND.
FT DISULFID 3111 3161
FT DISULFID 3120 3144
FT DISULFID 3136 3157
FT CARBOHYD 202 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2079 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2331 2331 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2558 2558 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2677 2677 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2861 2861 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3036 3036 (N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 32 236 MISSING (IN ISOFORM 2).
FT VARIANT 1679 1679 /FTID=VAR_001910.
FT VARIANT 2831 2831 D -> H.
FT CONFLICT 127 128 /FTID=VAR_001911.
FT CONFLICT 137 137 QS -> AK (IN REF. 6).
FT CONFLICT 2157 2157 R -> L (IN REF. 6).
FT CONFLICT 2157 2157 P -> R (IN REF. 4).
SQ SEQUENCE 3176 AA; 343548 MW; 5DF825632298B2DD CRC64;

```

Query Match 0.7%; Score 8; DB 1; Length 3176;

Best Local Similarity 100.0%; Pred. No. 68; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

202 IQGVLVOY 209

1676 IQGVLVOY 1683

```

RESULT 41
DMS2 AGAAN STANDARD; PRT; 72 AA.
AC 093224;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dermaseptin AA-3-3 precursor.
OS Agalychnis annae (Yellow-eye leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Phyllomedusinae; Agalychnis.
OC NCBI_TaxID=75990;
OX 1;
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RP TISSUE=Skin;
RC MEDLINE=98449786; PubMed=9774745;
RA Wechsclberger C.;
RT "Cloning of cDNAs encoding new peptides of the dermaseptin-family.";
RL Biochim. Biophys. Acta 1388:279-283 (1998).
CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-
CC POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING
CC MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SKIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005186; CAA06423.1; -
CC InterPro; IPR004275; Brevenin.

```

```

DR Pfam; PF03032; Brevenin; 1.
KW Antibiotic; Multigene family; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 41 POTENTIAL.
FT CHAIN 44 69 DERMASEPTIN AA-3-3.
FT PROPEP 71 72 POTENTIAL.
FT MOD_RES 69 69 AMIDATION (G-70 PROVIDE AMIDE GROUP)
SQ SEQUENCE 72 AA; 7996 MW; B16711501A43A42F CRC64;

```

1157 EEKREE 1163

23 EEKREE 29

```

RESULT 42
DMS2 AGAAN STANDARD; PRT; 73 AA.
AC 093222;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Dermaseptin AA-2-5 precursor.
OS Agalychnis annae (Yellow-eye leaf frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Phyllomedusinae; Agalychnis.
OC NCBI_TaxID=75990;
OX 1;
RN SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RP TISSUE=Skin;
RC MEDLINE=98449786; PubMed=9774745;
RA Wechsclberger C.;
RT "Cloning of cDNAs encoding new peptides of the dermaseptin-family.";
RL Biochim. Biophys. Acta 1388:279-283 (1998).
CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-
CC POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING
CC MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SKIN-SPECIFIC.
CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ005184; CAA06421.1; -
CC InterPro; IPR004275; Brevenin.
DR Pfam; PF03032; Brevenin; 1.
KW Antibiotic; Multigene family; Cleavage on pair of basic residues;
KW Amidation; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 42 POTENTIAL.
FT CHAIN 45 70 DERMASEPTIN AA-2-5.
FT PROPEP 72 73 POTENTIAL.
FT MOD_RES 70 70 AMIDATION (G-71 PROVIDE AMIDE GROUP)
SQ SEQUENCE 73 AA; 7855 MW; 4BDACDAD5DBF0944 CRC64;

```

1157 EEKREE 1163

DB 23 EEEKREE 29

RESULT 43

DMS3_PHYBI STANDARD; PRT; 74 AA.

AC P81485;

DR 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

OS Dermaseptin B11 precursor (Dermaseptin B3).

OC Phyllomedusa bicolor (Two-colored leaf frog).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae; Phyllomedusinae; Phyllomedusa.

CC NCBI_TaxId=8393;

OX [1]

RN SEQUENCE FROM N.A., SEQUENCE OF 46-73, AND MASS SPECTROMETRY.

RP TISSUE=Skin;

RC MEDLINE=98278974; PubMed=9614066;

RX Charpentier S., Antiche M., Mester J., Vouille V., Le Caer J.-P., Nicolas P., Delfour A., and molecular cloning of dermaseptins B, a family of skin peptide antibiotics."

RT "Structure, synthesis, and molecular cloning of dermaseptins B, a family of skin peptide antibiotics."

RL J. Biol. Chem. 273:14690-14697(1998).

CC -1- FUNCTION: POSSESSES A POTENT ANTIMICROBIAL ACTIVITY AGAINST GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA. PROBABLY ACTS BY DISTURBING MEMBRANE FUNCTIONS WITH ITS AMPHIPATHIC STRUCTURE.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MASS SPECTROMETRY: MW=2780.4; MW ERR=0.1; METHOD=Electrospray.

CC -1- SIMILARITY: BELONGS TO THE DERMASEPTIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation- the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Y16564; CAA16288.1; -

DR InterPro: IPR004275; Brevinin.

DR Pfam: PF03032; Brevinin; 1.

KW Antibiotic; Multigene family; Amphibian skin; Signal;

KV Cleavage on pair of basic residues.

FT SIGNAL 1 22 POTENTIAL.

FT PROPEP 23 43

FT CHAIN 46 74 DERMASEPTIN B11.

FT SEQUENCE 74 AA; 8254 MW; F9D5902A24F32C8D CRC64;

SO

Query Match 0.6%; Score 7; DB 1; Length 74;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EEEKREE 1163

DB 23 EEEKREE 29

RESULT 44

ACP_COMTE STANDARD; PRT; 77 AA.

AC P80318;

DR 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

OS Acyl carrier protein (ACP).

CC ACP.

CC Comamonas testosteroni (Pseudomonas testosteroni).

CC Bacteria; Proteobacteria; Beta subdivision; Comamonadaceae; Comamonas.

CC NCBI_TaxId=285;

CC (1)

RP SEQUENCE.

RC STRAIN=ATCC 11996;

RX MEDLINE=97315245; PubMed=9171419;

RA Tang L., Weisborn A.C., Kennedy E.P.;

RT "Domains of Escherichia coli acyl carrier protein important for membrane-derived-oligosaccharide biosynthesis."

RL J. Bacteriol. 179:3697-3705(1997).

CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid biosynthesis.

CC -1- PATHWAY: De novo fatty acid biosynthesis.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific serine of apo-ACP by acps. This modification is essential for activity because fatty acids are bound in thioester linkage to the sulfhydryl of the prosthetic group.

CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.

CC HSP, P02901; IACP.

DR InterPro: IPR003231; Acyl_carrier.

DR InterPro: IPR003880; Pantane_attach.

DR Pfam: PF00550; pd-binding; 1.

DR PRODOM: PD000887; Acyl_carrier; 1.

DR TIGRPFAMs: TIGR00517; acyl_carrier; 1.

DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.

DR PROSITE: PS00075; ACP_DOMAIN; 1.

KW Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine.

FT BINDING 36 36 PHOSPHOPANTHETHEINE.

SO SEQUENCE 77 AA; 8406 MW; 2954BE8D07944B71 CRC64;

QY 414 MALEDEF 420

DB 44 MALEDEF 50

RESULT 45

ACP_RALSO STANDARD; PRT; 79 AA.

AC O8Y0U1;

DR 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

OS Acyl carrier protein 1 (ACP 1).

GN ACP1 OR ACP OR RSC1053 OR RS04175.

OS Ralstonia solanacearum (Pseudomonas solanacearum).

CC Bacteria; Proteobacteria; Beta subdivision; Ralstonia group; Ralstonia.

CC NCBI_TaxId=305;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=GM11000;

RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Arlat M., Billault A., Broctier P., Gams J.C., Catolico L., Chandler M., Choisme N., Claudel-Renard C., Cunne S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T., Sigler P., Thebault P., Whalen M., Wincker P., Levy M., Weissenbach J., Boucher C.A.;

RA "Genome sequence of the plant pathogen Ralstonia solanacearum."

RL Nature 415:497-502(2002).

CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid biosynthesis (By similarity).

CC -1- PATHWAY: De novo fatty acid biosynthesis.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- PTM: 4'-phosphopantetheine is transferred from CoA to a specific serine of apo-ACP by acps. This modification is essential for activity because fatty acids are bound in thioester linkage to the sulfhydryl of the prosthetic group (By similarity).

CC -1- SIMILARITY: CONTAINS 1 ACYL CARRIER DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL646062; CAD14755.1; -
 DR InterPro: IPR003231; Acyl_carrier.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00550; pp-binding; 1.
 DR ProDom: PD000887; Acyl_carrier; 1.
 DR TIGRFAMs: TIGR00517; acyl_carrier; 1.
 DR PROSITE: PS50075; ACP DOMAIN; 1.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
 KM Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;
 KW Complete proteome.
 FT BINDING 37 37 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 79 AA; 8714 MW; F2737B0B5588E27B CRC64;
 Query Match 0.6%; Score 7; DB 1; Length 79;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 414 MALEDEF 420
 Db 45 MALEDEF 51
 RESULT 46
 AC XYLFA STANDARD; PRT; 79 AA.
 ID ACP_XYLFA
 AC 09PFI5;
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Acyl carrier protein (ACP).
 OS Xylella fastidiosa.
 GN ACP OR XF0672.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 NC NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme W., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins C.Y., Monteiro-Vitorello C.B.,
 RA Menck C.F.M., Mireca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Trai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago W.A., Zatz W., Weidman J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";

RL Nature 406:151-159(2000).
 CC -1- FUNCTION: Carrier of the growing fatty acid chain in fatty acid
 CC biosynthesis (By similarity).
 CC -1- PATHWAY: De novo fatty acid biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- PTM: 4-phosphopantetheine is transferred from CoA to a specific
 CC serine of apo-ACP by acps. This modification is essential for
 CC activity because fatty acids are bound in thioester linkage to the
 CC sulhydryl of the prosthetic group (By similarity).
 CC -1- SIMILARITY: CONTAINS 1 ACP CARRIER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE003911; AAF83482.1; ALT_INIT.
 DR HSSP: P02901; IACP.
 DR InterPro: IPR003231; Acyl_carrier.
 DR InterPro: IPR003880; Ppantne_attach.
 DR Pfam: PF00550; pp-binding; 1.
 DR ProDom: PD000887; Acyl_carrier; 1.
 DR TIGRFAMs: TIGR00517; acyl_carrier; 1.
 DR PROSITE: PS50075; ACP DOMAIN; 1.
 DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 1.
 KM Lipid synthesis; Fatty acid biosynthesis; Phosphopantetheine;
 KW Complete proteome.
 FT BINDING 37 37 PHOSPHOPANTETHEINE (BY SIMILARITY).
 SQ SEQUENCE 79 AA; 8782 MW; 4A58822134BEC6B CRC64;
 Query Match 0.6%; Score 7; DB 1; Length 79;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 414 MALEDEF 420
 Db 45 MALEDEF 51
 RESULT 47
 AC HAEBIN
 ID YAJC_HAEIN STANDARD; PRT; 97 AA.
 AC P4592;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein HI0241.
 GN HI0241.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 CC Haemophilus.
 NC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95550630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerevasch A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Usterback T.R., Hama M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE UPF0092 FAMILY. STRONG. TO E.COLI YAJC.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32710; AAC21909.1; -
DR TIGR; H10241; -
DR InterPro: IPR003849; Ya3C.
DR Pfam: PF02699; DUF219; 1.
DR TIGRFAMs: TIGR00739; Ya3C; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 5 25
FT POTENTIAL.
SQ SEQUENCE 97 AA; 10754 MW; 86F9A683A1D28358 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 97;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1055 LAKGTEV 1061
Db 42 LAKGTEV 48

RESULT 48
INS_CHICK STANDARD; PRT; 107 AA.
ID INS_CHICK
AC P01332;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Insulin precursor.
GN INS.
OS Gallus gallus (Chicken),
OS Meleagris gallopavo (Common turkey), and
OS Struthio camelus (Ostrich).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031, 9103, 8801;
RN 11
RP SEQUENCE FROM N.A.
RC SPECIES=Chicken;
RX MEDLINE=80222898; PubMed=7388949;
RA Perler F., Efferatiadis A., Lomedico P., Gilbert W., Kolodner R.,
RA Dodgson J.B.;
RT "The evolution of genes: the chicken preproinsulin gene";
RT Cell 20:555-566(1980).
RN 12
RP SEQUENCE FROM N.A.
RC SPECIES=Chicken; TISSUE=pancreas;
RX Hasagawa S., Honda K., Naka K., Yonekura H., Okamoto H., Hikami Y.,
RT "Isolation of a cDNA encoding chicken insulin precursor";
RT Anim. Sci. Technol. 62:867-869(1991).
RN 13
RP SEQUENCE OF 25-54 AND 87-107.
RC SPECIES=Chicken;
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin";
RT Am. J. Med. 40:662-666(1966).
RN 14
RP SEQUENCE OF 25-54 AND 87-107.
RC SPECIES=M.gallopavo;
RX MEDLINE=72259992; PubMed=5066110;
RA Jentsch J.;
RT "Structure and increased activity of insulin from the turkey
RT (Meleagris gallopavo).";
RT Hope-Seyler's Z. Physiol. Chem. 353:980-986(1972).
RN 15
RP SEQUENCE OF 25-54 AND 87-107.

RC SPECIES=S.camelus;
RX MEDLINE=8814456; PubMed=3045031;
RA Evans T.K., Lithauer D., Oelofsen W.;
RT "Purification and primary structure of ostrich insulin";
RT Int. J. Pept. Protein Res. 31:454-462(1988).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE PARTIAL SEQUENCE OF TURKEY AND OSTRICH ARE
CC IDENTICAL WITH THAT OF CHICKEN.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00874; AAA48925.1; -
DR EMBL; J00872; AAA48925.1; JOINED.
DR EMBL; V00416; CAA23707.1; -
DR EMBL; V00418; CAA23708.1; -
DR EMBL; X58993; CAA1738.1; -
DR PIR; A01598; IPCH.
DR PIR; A01599; INTR.
DR PIR; JK0017; INOS.
DR HSSP; P01308; IHS.
DR InterPro: IPR004825; Ins/IGF/relax.
DR Pfam: PF00049; Insulin.1.
DR PRINTS; PR00276; INSULIN.1.
DR PRINTS; PR00277; INSULIN.
DR SMART; SM00078; IIGF.1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 1 25
FT PROPEP 57 84
FT CHAIN 87 107
FT DISULFID 31 93
FT DISULFID 43 106
FT DISULFID 92 97.
SQ SEQUENCE 107 AA; 11981 MW; 3D43C5D82B25DCDD CRC64;

Query Match 0.6%; Score 7; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1136 LIALIVF 1142
Db 10 LIALIVF 16

RESULT 49
HDEA_ECOLI STANDARD; PRT; 110 AA.
ID HDEA_ECOLI
AC P26604;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein hdea precursor (10k-S protein).
GN HDEA OR B3510 OR Z4922 OR ECS4390.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN 11

RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-40.
 RC STRAIN=K12;
 RX MEDLINE=93204884; PubMed=8455549;
 RA Yoshida T., Ueguchi C., Yamada H., Mizuno T.;
 RT "Function of the Escherichia coli nucleoid protein, H-NS: molecular
 RT analysis of a subset of proteins whose expression is enhanced in a
 RT hns deletion mutant.";
 RL Mol. Gen. Genet. 237:113-122(1993).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes.";
 RL Nucleic Acids Res. 22:2576-2586(1994).
 RN [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobo T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [15]
 RP GENE NAME.
 RX MEDLINE=94064579; PubMed=8244952;
 RA Yoshida T., Ueguchi C., Mizuno T.;
 RT "Physical map location of a set of Escherichia coli genes (hde) whose
 RT expression is affected by the nucleoid protein H-NS.";
 RL J. Bacteriol. 175:7747-7748(1993).
 RN [16]
 RP SEQUENCE OF 22-41.
 RC STRAIN=K12 / W3110;
 RX Pasquali C., Sanchez J.-C., Ravier F., Golaz O., Hughes G.J.,
 RA Fruiter S., Paquet N., Wilkins M., Appel R.D., Baloch A.,
 RA Hochstrasser D.F.;
 RT Submitted (SEP-1994) to the SWISS-PROT data bank.
 RN [17]
 RP SEQUENCE OF 22-33.
 RC STRAIN=K12 / EMG2;
 RX MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robison K., Church G.M.;
 RT "Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN [18]
 RP SEQUENCE OF 22-31.
 RC STRAIN=K12;
 RX MEDLINE=99065675; PubMed=9868784;
 RA Wasinger V.C., Humphrey-Smith I.;
 RT "Small genes/gene-products in Escherichia coli K-12.";
 RL FEMS Microbiol. Lett. 169:375-382(1998).
 RN [19]
 RP X-RAY CRYSTALLOGRAPHY
 RX MEDLINE=98400493; PubMed=9731767;
 RA Yang F., Gustafson K.R., Boyd M.R., Wlodawer A.;
 RT "Crystal structure of Escherichia coli HdeA.";

RL Nat. Struct. Biol. 5:763-764(1998).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=20090957; PubMed=10623550;
 RA Gajwala K.S., Burley S.K.;
 RT "HdeA, a periplasmic protein that supports acid resistance in
 RT pathogenic enteric bacteria.";
 RL J. Mol. Biol. 295:605-612(2000).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, D11109; BAA01883.1; -;
 CC EMBL, U00039; AAB18486.1; -;
 CC EMBL, AE000427; AAC76535.1; -;
 CC EMBL, AE005576; AAG58651.1; -;
 CC EMBL, AP002565; BAB37813.1; -;
 CC PIR, S30266; S30266.
 CC PIR, S30268; S30268.
 CC PDB, 1B68; 16-SEP-98.
 CC PDB, 1D48; 10-DEC-99.
 CC SWISS-2DPAGE; P26604; COLI.
 CC Ecogene; Ecol198; hdeA.
 CC Periplasmic; Signal; 3D-structure; Complete proteome.
 FT SIGNAL 1 21
 FT CHAIN 22 110 PROTEIN HDEA.
 FT DISULFID 39 87
 SQ SEQUENCE 110 AA; 11858 MW; 063262C4863FA2E9 CRC64;
 QY 1131 IGGILLL 1137
 Db 9 LGGILLL 15
 RESULT 50
 IGF MYXGL
 ID IGF MYXGL STANDARD; PRT; 139 AA.
 AC P22618;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Insulin-like growth factor precursor (IGF) (Fragment).
 OS Myxine glutinosa (Atlantic hagfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
 OC Myxiniidae; Myxiniinae; Myxine.
 OC NCBI_Taxid=7769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91115860; PubMed=1989990;
 RA Nagamatsu S., Chan S.J., Falkner S., Steiner D.F.;
 RT "Evolution of the insulin gene superfamily. Sequence of a
 RT preproinsulin-like growth factor cDNA from the Atlantic hagfish.";
 RL J. Biol. Chem. 266:2397-2402(1991).
 CC -1- FUNCTION: THE INSULIN-LIKE GROWTH FACTORS, ISOLATED FROM PLASMA,
 CC ARE STRUCTURALLY AND FUNCTIONALLY RELATED TO INSULIN BUT HAVE A
 CC MUCH HIGHER GROWTH-PROMOTING ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; M57735; AAA49265.1; -.	
DR	PIR; A38612; A38612.	
DR	HSSP; P01344; IGF2.	
DR	InterPro; IPR004825; Ins/IGF/relax.	
DR	Pfam; PF00049; Insulin; 1.	
DR	SMART; SMO0078; IIGF; 1.	
DR	PROSITE; PS00262; INSULIN; 1.	
KW	Insulin family; Mitogen; Growth factor; Signal.	
FT	NON_TER	1
FT	SIGNAL	38
FT	CHAIN	39
FT	DOMAIN	39 139
FT	DOMAIN	39 67
FT	DOMAIN	82
FT	DOMAIN	83
FT	DOMAIN	104
FT	DOMAIN	104 113
FT	DOMAIN	114
FT	DOMAIN	139
SO	SEQUENCE	139 AA; 16067 MW; 2FC886C8D074FAc1 CRC64;

```

CC RESULT 51
CCD YEAST ID_CDD YEAST STANDARD; PRT; 142 AA.
AC 006549;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytidine deaminase (EC 3.5.4.5) (Cytidine aminohydrolase) (CDA)
GN CDD1 OR YLR245C OR L672.13.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetaceae.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=FL100;
RX MEDLINE=99431800; PubMed=10501935.
RA Kurtz J.-E., Exinger F., Erbs P., Jund R.;
RT "New insights into the pyrimidine salvage pathway of Saccharomyces
RT cerevisiae: requirement of six genes for cytidine metabolism.";
RL Curr. Genet. 36:130-136(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Knudsen T., Hallsworth K., Hawkins J., Hillier L., Jier T.,
RA Johnston D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan N., Pauley A., Peluso D.,
RA Ritken L., Riles L., Taich A., Tsvetakis E., Vignati D.,
RA Milcox L., Mohlman P., Vaundin M., Wilson R., Waterson R.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIS ENZYME SCAVENGE EXOGENOUS AND ENDOGENOUS CYTIDINE
CC -1- AND 2-DEOXYCYTIDINE FOR UMP SYNTHESIS.
CC -1- CATALYTIC ACTIVITY: Cytidine + H(2)O = uridine + NH(3) .
CC -1- COFACTOR: ZINC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

SO	SEQUENCE	142 AA;	1536 MW;	9622BDB9CCBB30D51F	CRC64
DR	EMBL; U080089; AAD04031.1; -.				
DR	EMBL; U08065; AAB67399.1; -.				
DR	HSSP; P13652; 1ALN.				
DR	SGD; S0004235; CDD1.				
DR	InterPro; IPR002125; dCMP_cyt_deam.				
DR	Pfam; PF00383; dCMP_cyt_deam.1.				
DR	PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.				
KM	Hydrolase; zinc.				
FT	METAL	61		ZINC (BY SIMILARITY).	
FT	METAL	96		ZINC (BY SIMILARITY).	
FT	METAL	99		ZINC (BY SIMILARITY).	
SO	SEQUENCE	142 AA;	1536 MW;	9622BDB9CCBB30D51F	CRC64

RESULT 52			
MK_CHICK	STANDARD;	PRT;	142 AA.
ID	PK2052;		
DT	01-MAR-1992 (Rel. 21, Created)		
DT	01-MAR-1992 (Rel. 21, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	Mdkline precursor (Retinoic acid-induced heparin-binding protein) (R1-HB).		
GN	R1HB.		
OS	Gallus gallus (Chicken).		
OC	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Archaeoptila; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
OX	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 80-85; 112-118 AND 121-133.		
RC	TISSUE=Embryo;		
RA	MEDLINE=91207359; PubMed=2018506;		
RA	Utrios P., Duprez D., Je Caer J.-P., Courtois Y., Vigny M., Laurent M.;		
RT	"Molecular cloning of R1-HB, a heparin binding protein regulated by		
RT	retinoic acid.";		
RL	Biochem. Biophys. Res. Commun. 175:617-624(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=95010085; PubMed=7925417;		
RT	Duprez D., Treasger J., Pecqueur C., Vigny M.R.;		
RT	"Organisation and promoter activity of the		
RT	retinoic-acid-induced-heparin-binding (R1HB) gene.";		
RL	Eur. J. Biochem. 224:931-941(1994).		
RN	[3]		
RP	SEQUENCE OF 22-77.		
RA	MEDLINE=91128406; PubMed=1993066;		
RA	Raulais D., Lagente-Chevallier O., Guettet C., Duprez D., Courtois Y.,		
RA	Vigny M.;		
RT	"A new heparin binding protein regulated by retinoic acid from chick		
RT	embryo.";		
RL	Biochem. Biophys. Res. Commun. 174:708-715(1991).		
RN	[4]		
RP	SEQUENCE OF 22-77.		
RC	SMRAIN=White leghorn; TISSUE=Embryo;		
RA	MEDLINE=90108010; PubMed=2558016;		
RA	Vigny M., Raulais D., Puzenat N., Duprez D., Hartman M.P.,		
RA	Jeanny J.C., Courtois Y.;		
RT	"Identification of a new heparin-binding protein localized within		
RT	chick basement membranes.";		

```

RL Eur. J. Biochem. 186:733-740(1989).
CC -1- FUNCTION: HAS MITOGENIC ACTIVITY, AND NEURITE EXTENSION ACTIVITY
CC FOR PC12 CELLS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES IN EARLY EMBRYONIC
CC TISSUES, AND CELL SURFACE OF NEUROECTODERMAL CELLS.
CC -1- DEVELOPMENTAL STAGE: ESSENTIALLY EXPRESSED DURING EMBRYOGENESIS.
CC -1- INDUCTION: BY RETINOIC ACID.
CC -1- SIMILARITY: BELONGS TO THE PLEIOTROPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M61754; -; NOT ANNOTATED_CDS.
CC DR EMBL; X76482; CAAS4020.1; -.
CC DR PIR; J70573; J70573.
CC DR HSSP; P21741; 1MKC.
CC DR InterPro; IPR000762; PTN_MK.
CC DR Pfam; PFO1091; PTN_MK.1.
CC DR PRINTS; PR00269; PTNMDKINE.
CC DR ProDom; PD005592; PTN_MK.1.
CC DR SMART; SM00193; PTN.1.
CC DR PROSITE; PS00619; PTN_MK.1; 1.
CC DR PROSITE; PS00620; PTN_MK.2; 1.
CC KW Growth factor; Mitogen; Differentiation; Heparin-binding; Signal.
CC FT SIGNAL 1 21
CC FT CHAIN 22 142 MDKINE.
CC FT DISULFID 36 60 BY SIMILARITY.
CC FT DISULFID 44 69 BY SIMILARITY.
CC FT DISULFID 51 73 BY SIMILARITY.
CC FT DISULFID 83 115 BY SIMILARITY.
CC FT DISULFID 93 125 BY SIMILARITY.
CC FT CONFLICT 89 89 S -> R (IN REF. 2).
CC FT CONFLICT 93 93 C -> G (IN REF. 2).
CC SQ SEQUENCE 142 AA; 15579 MW; 9D05CAF89558451B CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLLLL 1140
DB 7 LLLLLL 13

RESULT 53
GLB_APLJU STANDARD; PRT; 144 AA.
AC P14393;
ID GLB_APLJU
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Globin (Myoglobin).
OS Aplysia juliana (Sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspeidae;
OC Aplysiidae; Aplysia.
OC NCBI_TaxID=6506;
OX NCBI_TaxID=6506;
RN (1)
RP SEQUENCE.
RX MEDLINE=85134891; PubMed=6527386;
RT Takagi T., Iida S., Matsuo K., Shikama K.;
RT "Aplysia myoglobins with an unusual amino acid sequence.";
RL J. Mol. Biol. 180:1179-1184(1984).
CC -1- SUBUNIT: Monomer.
CC -1- MISCELLANEOUS: THIS MOLUSCAN GLOBIN LACKS ONE OF THE HEME-BINDING
CC HISTIDINE RESIDUES FOUND IN MOST OTHER GLOBINS.
CC -1- SIMILARITY: BELONGS TO THE GLOBIN FAMILY.
CC DR PIR; A44585; A44585.
CC DR HSSP; P02210; 2FAL.

```

```

DR InterPro; IPR002336; Erythecurin.
DR InterPro; IPR000971; Globin.
DR Pfam; PF00042; globin.1.
DR PRINTS; PR00611; ERYTHRUORIN.
DR PROSITE; PS01033; GLOBIN.1.
KW Heme; Oxygen transport; Transport; Muscle; Acetylation.
FT MOD RES 1 1 ACETYLATION.
FT METIL 95 95 IRON (HEME PROXIMAL LIGAND)
FT (BY SIMILARITY).
SQ SEQUENCE 144 AA; 15016 MW; 28FCF0FC378E50FB CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 984 GLIISAL 990
DB 133 GLIISAL 139

RESULT 54
RHIC_RHILV STANDARD; PRT; 149 AA.
AC Q03315;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein rhic precursor.
GN RHIC.
OS Rhizobium leguminosarum (biovar viciae).
OG Plasmid sym pRL101.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=387;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283758; PubMed=1597418;
RA Cubo M.T., Economou A., Murphy G.J., Johnston A.W., Downie J.A.;
RT "Molecular characterization and regulation of the
RT rhizosphere-expressed genes rhiaBCR that can influence nodulation by
RT Rhizobium leguminosarum biovar viciae.";
RL J. Bacteriol. 174:4026-4035(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN PLANT-MICROBE INTERACTION.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M98835; AAA26359.1; -.
CC DR PIR; C41887; C41887.
KW Plasmid; Periplasmic; Signal.
FT SIGNAL 1 23
FT CHAIN 24 149 POTENTIAL.
FT PROTEIN RHIC.
SQ SEQUENCE 149 AA; 15299 MW; F41A72AB1646B193 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 754 ALTVTFA 760
DB 15 ALTVTFA 21

RESULT 55
BIK_HUMAN STANDARD; PRT; 160 AA.
AC Q13523; Q16582;

```

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Bcl-2 interacting killer (Apoptosis inducer NBK) (BP4) (BIP1).
 GN BIK OR NBK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=B-cell.
 RX MEDLINE=96068922; PubMed=7478623;
 RA Boyd J.M., Gallo G.J., Elangovan B., Houghton A.B., Malstrom S.,
 RA Avery B.J., Ebb R.G., Subramanian T., Chittenden T., Lutz R.J.,
 RA Chinnadurai G.;
 RT "Bik, a novel death-inducing protein shares a distinct sequence motif
 RT with Bcl-2 family proteins and interacts with viral and cellular
 RT survival-promoting proteins.";
 RT Oncogene 11:1921-1928(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96413338; PubMed=8816500;
 RA Han J., Sabbatini P., White E.;
 RT "Induction of apoptosis by human NBK/Bik, a BHL-2-containing protein
 RT that interacts with Bcl-2.";
 RT Mol. Cell. Biol. 16:5857-5864(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymphoid;
 RA Pun K.-T., Farrow S.N., Raven T., Wride C.J., White J.H.M., Brown R.;
 RT Submitted (Jun-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99431905; PubMed=10500065;
 RA Casellas A., Ino Y., Louis D.N., Ramesh V., Gusella J.F., Rustgi A.K.;
 RT "Mapping of a target region of allelic loss to a 0.5-cM interval on
 RT chromosome 22q13 in human colorectal cancer.";
 RT Gastroenterology 117:831-837(1999).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Buskiewicz R., Beare D.M.,
 RA Clump M., Smith L.J., Ainscough R., Almeida J.P., Babbage A.K.,
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
 RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cordy N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
 RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall S.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,
 RA McElay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.C.T.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Senter H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spurgeon L., Steward C.A., Sultson J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshitaki Y., Aoki N., Mitsuyma S.,
 RA Roe B.A., Chen F., Chu Y., Hu P., Hua A., Kenton S., Do A., Do T.,
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren O., Salaj E., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,
 RA Zhan W., Zhang G., Chisoe S., Murray J., Miller N., Wink P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,

RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latschke P., Layman D., Ozersky P., Rohlfing T.,
 RA Schick P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korfi I., Bedell J.A., Hillier L., Mardis E., Weststrom R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saita S., Budarf M.L.,
 RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dunamski J.P., Peyrard M., Kedra D.,
 RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkinson P., Bodenreich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RT Nature 402:489-495(1999).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strausberg R.;
 RT Submitted (Jan-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP MUTAGENESIS, AND FUNCTION OF BHL-2 DOMAIN.
 RX MEDLINE=96091131; PubMed=8521816;
 RA Chittenden T., Flemington C., Houghton A.B., Ebb R.G., Gallo G.J.,
 RA Elangovan B., Chinnadurai G., Lutz R.J.;
 RT "A conserved domain in Bak, distinct from Bhl and Bh2, mediates cell
 RT death and protein binding functions.";
 RT EMBO J. 14:5589-5596(1995).
 CC -1- FUNCTION: ACCELERATES PROGRAMMED CELL DEATH. BINDING TO THE
 CC APOPTOSIS REPRESSORS BCL-X(L), BHRF1, BCL-2 OR ITS ANOGENOUS
 CC HOMOLOG BHL-19K PROTEIN SUPPRESSES THIS DEATH-PROMOTING ACTIVITY.
 CC DOES NOT INTERACT WITH BAX.
 CC -1- SUBCELLULAR LOCATION: AROUND THE NUCLEAR ENVELOPE. AND IN
 CC CYTOPLASMIC MEMBRANES.
 CC -1- DOMAIN: INTACT BHL-2 DOMAIN IS REQUIRED BY BIK, BID, BAK, BAD AND
 CC BAX FOR THEIR PRO-APOPTOTIC ACTIVITY AND FOR THEIR INTERACTION
 CC WITH ANTI-APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.
 CC APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG 3 (BHL) DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U34584; AAC50413.1; -
 DR EMBL, U49730; AAC79124.1; -
 DR EMBL, X89986; CAA62013.1; -
 DR EMBL, AF174424; AAF01156.1; -
 DR EMBL, AF174421; AAF01156.1; JOINED.
 DR EMBL, AF174422; AAF01156.1; JOINED.
 DR EMBL, AF174423; AAF01156.1; JOINED.
 DR EMBL, AL022237; CAA18260.2; -
 DR EMBL, BC001599; AAH01599.1; -
 DR Genew; HGNC:1051; BIK.
 DR MIM: 603392; -
 DR InterPro: IPR000712; Bcl2_BH.
 DR PROSITE: PS01259; BH3; 1.
 KW Apoptosis; Transmembrane.
 FT DOMAIN 57 71 BH3.
 FT TRANSMEM 136 156 POTENTIAL.
 FT DOMAIN 137 158 LEUCINE-ZIPPER (POTENTIAL).
 FT CONFLICT 149 150 PL->LP (IN REF. 1).
 SQ SEQUENCE 160 AA; 18016 MW; 890344443F5A136 CRC64;
 Query Match 0.6%; Score 7; DB 1; Length 160;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1134 LLLALL 1140
 DB 141 LLLALL 147

```

RESULT 56
YCRS MYXXA STANDARD; PRT; 170 AA.
ID YCRS MYXXA
AC 006929;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 18.1 kDa protein in CARS 3'region (ORF).
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cytoobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK101;
RX MEDLINE=95020544; PubMed=7934835;
RA McGowan S.J., Gorham H.C., Hodgson D.A.;
RT "Light-induced carotenogenesis in Myxococcus xanthus: DNA sequence
RT analysis of the car region.";
RL Mol. Microbiol. 10:713-735(1993).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; X71062; CAA50384.1; -
CC DR PIR; S33152; S33152.
CC DR PIR; S39880; S39880.
CC DR PIR; S35742; S35742.
CC KW Hypothetical protein.
CC SQ SEQUENCE 170 AA; 18052 MW; E446D38979EBF02D CRC64;
-----
Query Match
Best Local Similarity 0.6%; Score 7; DB 1; Length 170;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 403 LEGHRL 409
DB 22 LEGHRL 28
-----
RESULT 57
ATPF SYNPE STANDARD; PRT; 171 AA.
ID ATPF SYNPE
AC P08447;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase B chain (EC 3.6.3.14) (Subunit I).
GN ATPF.
OS Synecococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1139;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=87311713; PubMed=3041005;
RA Cozens A.L., Walker J.E.;
RT "The organization and sequence of the genes for ATP synthase subunits
RT in the cyanobacterium Synecococcus 6301. Support for an
RT endosymbiotic origin of chloroplasts.";
RL J. Mol. Biol. 194:359-383(1987).
-----
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE B CHAIN FAMILY.

```

```

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; X05102; CAA28926.1; -
CC DR PIR; S10829; LMVCL.
CC DR InterPro; IPR002146; ATPsynt_B/B' sub.
CC DR Pfam; PF00430; ATP-synt_B; 1.
CC KW Hydrogen ion transport; Transmembrane; CF(0).
CC FT TRANSMEM 24 40
CC FT POTENTIAL.
CC SQ SEQUENCE 171 AA; 18681 MW; ADB55F8FA6D98C2C CRC64;
-----
Query Match
Best Local Similarity 0.6%; Score 7; DB 1; Length 171;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 266 EARLLV 272
DB 86 EARLLV 92
-----
RESULT 58
YF87 METJA STANDARD; PRT; 171 AA.
ID YF87 METJA
AC Q58982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1587.
GN MJ1587.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodex A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Furumann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
-----
CC -1- SIMILARITY: TO M.JANNASCHII MJ0417.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; U67589; AAB99616.1; -
CC DR TIGR; MJ1587; -
CC DR InterPro; IPR004013; PNP_C.
CC DR InterPro; IPR003141; PNP_N.
CC DR Pfam; PF02231; PNP_N; 1.
CC DR Pfam; PF02811; PNP_C; 1.
CC DR SMART; SM00481; POLIITAC; 1.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 171 AA; 19198 MW; 7D4F7B936C5C7472 CRC64;

```

Query Match 0.6%; Score 7; DB 1; Length 171;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 ALGDRIF 355
 |||||
 DB 123 ALGDRIF 129

RESULT 59

OC PORC_METUA STANDARD; PRT; 178 AA.
 ID PORC_METUA
 AC 057717;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pyruvate synthase subunit PORC (EC 1.2.7.1). (Pyruvate oxidoreductase gamma chain) (POR) (Pyruvic-ferredoxin oxidoreductase gamma subunit).
 GN PORC OR MJO269.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.
 OK NCBI_TaxID=2190;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Keriakova A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.";
 RL Science 273:1058-1073 (1996).
 CC -1- CATALYTIC ACTIVITY: Pyruvate + CoA + oxidized ferredoxin = acetyl-
 CC -1- SUBUNIT: HETEROTRIMER OF ONE ALPHA, ONE BETA, ONE DELTA AND ONE
 CC GAMMA CHAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U67482; AAB89256.1; -
 DR TIGR: MJO269; -
 DR InterPro: IPR002869; POR.
 DR Pfam: PF01558; POR; 1.
 KW Oxidoreductase; Complete proteome.
 SQ SEQUENCE 178 AA; 19202 MW; 3758E7E7C70FAE61 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 178;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 465 LKKDGA 471
 |||||
 DB 87 LKKDGA 93

RESULT 60
 HYFH_ECOLI STANDARD; PRT; 181 AA.
 ID HYFH_ECOLI
 AC F77423;
 DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hydrogenase-4 component H.
 GN HYFH OR B2488.

OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OK NCBI_TaxID=562;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Andrews S.C., McClay J., Ambler A., Quail M., Berts B.C., Guest J.R.,
 RA Submitted (Oct-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1233-1244 (1997).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubraman S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113 (1997).
 CC -1- FUNCTION: PROBABLE ELECTRON TRANSFER PROTEIN FOR HYDROGENASE 4.
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
 CC 'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M61654; AAB88570.1; -
 DR EMBL: AE000335; AAC75541.1; -
 DR EMBL: D90877; BAA16376.1; -
 DR HSPB; P00195; ICLP.
 DR Ecogene; EGI4216; hyfH.
 DR InterPro: IPR001450; 4Fe4S_Ferredoxin.
 DR Pfam: PF00037; Fe4; 2.
 DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 2.
 KW Electon transport; 4Fe-4S; Iron-sulfur; Complete proteome.
 SQ SEQUENCE 181 AA; 20159 MW; 5DBBA3C7C7EBD85 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 181;
 Best Local Similarity 100.0%; Pred. No. 68;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 TDVLVA 507

Db 171 TDVLVA 177

RESULT 61
UVP1_ECOLI STANDARD; PRT; 198 AA.

AC P18957;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)

GN UVP1 protein.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Escherichia

OX NCBI_TaxID=562;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=89384434; PubMed=2550763;

RX Gigliani F., Sporeno E., Perri S., Battaglia P.A.;

RT "The uvp1 gene of plasmid pr cooperates with mucAB genes in the DNA

repair process.";

RL Mol. Gen. Genet. 218:18-24(1989).

CC -1- FUNCTION: COOPERATES WITH MUCAB GENES IN THE DNA REPAIR PROCESS.

CC IT COULD BE A RESOLVASE-INVERTASE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE SITE-SPECIFIC RECOMBINASE RESOLVASE

FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X16119; CAA34249.1; -

DR PIR; S04903; S04903.

DR HSSP; P03012; 2RSL.

DR InterPro; IPR001822; Recombinase.

DR Pfam; PF00239; Resolvase; 1.

DR PROSITE; PS00397; HTH_7; 1.

DR PROSITE; PS00397; RECOMBINASES_1; FALSE_NEG.

DR PROSITE; PS00398; RECOMBINASES_2; 1.

KW DNA repair; Plasmid; DNA-binding; DNA recombination; DNA integration;

KM DNA invertase.

FT ACT_SITE 9

FT TRANSIENT COVALENT LINKAGE TO DNA DURING

STRAND CLEAVAGE AND REJOINING

(BY SIMILARITY).

FT DNA BIND 168 187

FT H-T-H MOTIF (BY SIMILARITY).

SO SEQUENCE 198 AA; 22127 MW; 8011E768EB0548 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 198;

Best Local Similarity 100.0%; Pred. No. 74;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1003 SUSQVIT 1009

Db 74 SUSQVIT 80

RESULT 62

RACI_DICDI STANDARD; PRT; 205 AA.

AC Q9GPR2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE PAS-related protein rac1.

GN RAC1.

OS Dictyostelium discoideum (Slime mold).

OC Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

OX NCBI_TaxID=44689;

RN (1)

RP SEQUENCE FROM N.A.

RX STRAIN=AX4;

RA MEDLINE=21127961; PubMed=11222756;

RT Rivero F., Dislich H., Glockner G., Noegel A.A.;

RL "The Dictyostelium discoideum family of Rho-related proteins.";

Nucleic Acids Res. 29:1068-1079(2001).

CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AF310895; AAG45138.1; -

DR HSSP; P5154; 1B96.

DR DictyDB; DD27272; rac1.

DR InterPro; IPR003578; GTPase_Rho.

DR InterPro; IPR001230; Prenyl_site.

DR InterPro; IPR001806; Ras_trnstrng.

DR Pfam; PF00071; ras; 1.

DR PRINTS; PR00449; RASFRNFRMNG.

DR SMART; SM00174; RHO; 1.

DR TIGRfams; TIGR00231; small_GTP_1.

KW GTP-binding; Prenylation; Lipoprotein.

FT NP BIND 12 19

FT GTP (BY SIMILARITY).

FT NP BIND 59 63

FT GTP (BY SIMILARITY).

FT NP BIND 119 122

FT GTP (BY SIMILARITY).

FT DOMAIN 34 42

FT EFFECTOR REGION (POTENTIAL).

FT LIPID 202 202

FT GERANYL-GERANYL (BY SIMILARITY).

FT LIPID 202 202

FT GERANYL-GERANYL (BY SIMILARITY).

SO SEQUENCE 205 AA; 22997 MW; F9EFD31576C45CC CRC64;

Query Match 0.6%; Score 7; DB 1; Length 205;

Best Local Similarity 100.0%; Pred. No. 76;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1046 QVVRCHL 1052

Db 176 QVVRCHL 182

RESULT 63

GPBB_MOUSE STANDARD; PRT; 206 AA.

AC P56400;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Platelet glycoprotein Ib beta chain precursor (GP-Ib beta) (GP1bB)

DE (GP1b-Beta).

GN GP1bB.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=97403789; PubMed=9259114;

RA Kiteguchi T., Murata M., Anbo H., Moriki T., Ikeda Y.;

RT "Characterization of the gene encoding mouse platelet glycoprotein Ib

beta.";

Thromb. Res. 87:235-244(1997).

CC -1- FUNCTION: GP-IB, A SURFACE MEMBRANE PROTEIN OF PLATELETS.

PARTICIPATES IN THE FORMATION OF PLATELET PLUGS BY BINDING TO VON

WILLEBRAND FACTOR, WHICH IS ALREADY BOUND TO THE SUBENDOTHELIUM

(BY SIMILARITY).

CC -1- SUBUNIT: GP-IB ALPHA AND BETA ARE DISULFIDE LINKED. GP-IX IS

CC COMPLEXED WITH THE GP-IB HETERODIMER VIA A NON COVALENT LINKAGE
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: PLATELET ACTIVATION APPARENTLY INVOLVES DISRUPTION
 CC OF THE MACROMOLECULAR COMPLEX OF GP-IB WITH THE PLATELET
 CC GLYCOPROTEIN IX (GP-IX) AND DISSOCIATION OF GP-IB FROM THE
 CC ACTIN-BINDING PROTEIN.
 CC -1- SIMILARITY: CONTAINS 1 LEUCINE-RICH REPEAT (LRR).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB001419; BAA22424.1; -.
 CC MGD; MGI:107852; GpIb.
 CC InterPro; IPR001611; LRR.
 CC InterPro; IPR000483; LRR_Cterm.
 CC InterPro; IPR000372; LRR_Nterm.
 CC Pfam; PF00560; LRR_1.
 CC Pfam; PF01462; LRRNT; 1.
 CC Pfam; PF01463; LRRCT; 1.
 CC SMART; SM00082; LRRCT; 1.
 CC SMART; SM00013; LRRNT; 1.
 CC Platelet; Transmembrane; Glycoprotein; Hemostasis; Blood coagulation;
 CC Signal; Phosphorylation; Cell adhesion; Leucine-rich repeat.
 CC SIGNAL
 CC CHAIN 1 26
 CC FT CHAIN 27 206
 CC FT DOMAIN 147 172
 CC FT TRASMEN 148 172
 CC FT DOMAIN 173 206
 CC FT REPEAT 60 83
 CC FT CAROXYD 66 66
 CC FT MOD RES 191 191
 CC SQ SEQUENCE 206 AA; 21762 MW; AC4BCB4DFA226F1D CRC64;
 CC
 CC Query Match
 CC Best Local Similarity 100.0%; Score 7; DB 1; Length 206;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1134 LLLALL 1140
 CC Db 13 LLLALL 19
 CC
 CC RESULT 64
 CC RL13 SCHPO STANDARD; PRT; 208 AA.
 CC ID 074175;
 CC AC 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE 60S ribosomal protein L13.
 CC GN RPL13 OR SPAC664.05.
 CC OS Schizosaccharomyces pombe (fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 CC RX NCB1
 CC RP SEQUENCE FROM N.A.
 CC RT "S. pombe ribosomal protein L13 homolog."
 CC RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=972;
 CC RX MEDLINE=21648401; PubMed=11859360;
 CC Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 CC Sgourou J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jags K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean C.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Roben J., Grymopre B.,
 RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel K., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer J., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambolt R., Punnett B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas R., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
 RA "The genome sequence of Schizosaccharomyces pombe."
 RA Nature 415:871-880(2002).
 RL Nature 415:871-880(2002).
 CC -1- SIMILARITY: BELONGS TO THE L13E FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB016216; BAA1740.1; -.
 CC EMBL; AL136235; CAB65806.1; -.
 CC InterPro; IPR001380; Ribosomal_L13E.
 CC Pfam; PF01294; Ribosomal_L13e; 1.
 CC Prodom; PD004443; Ribosomal_L13E; 1.
 CC DR PROSITE; PS01104; RIBOSOMAL_L13E; 1.
 CC RX RIBOSOMAL
 CC SQ SEQUENCE 208 AA; 23529 MW; B98EF67308F53F45 CRC64;
 CC
 CC Query Match
 CC Best Local Similarity 100.0%; Score 7; DB 1; Length 208;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 709 AGARAF 715
 CC Db 189 AGARAF 195
 CC
 CC RESULT 65
 CC RPSD MYCTU STANDARD; PRT; 212 AA.
 CC ID 050712;
 CC AC 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Probable RNA polymerase sigma-D factor.
 CC GN SIGD OR RV3414C OR MTCY78.15.
 CC OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteriaceae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 CC NCBI_TaxID=1773;
 CC RX NCB1
 CC RP SEQUENCE FROM N.A.
 CC RT STRAIN=H37RV;
 CC RC MEDLINE=98295987; PubMed=9634230;
 CC Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gao S., Barry C.E., III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulem J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
 CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
 CC THEN IS RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SIGMA-70 FACTOR FAMILY. ECF SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 277165; CAB01009.1; -
 CC EMBL; AE007157; AAK47861.1; ALT_INIT.
 CC TIGR; MT3523; -
 CC
 CC DR EMBL; 277165; CAB01009.1; -
 CC DR TIGR; MT3523; -
 CC
 CC DR Tuberculinf; Rv3414c; -
 CC DR InterPro; IPR000838; Sigma70_ECF.
 CC DR Pfam; PF00776; Sigma70_ECF; 1.
 CC DR PROSITE; PS01063; SIGMA70_ECF; 1.
 CC KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
 CC KM DNA-binding; Complete proteome.
 CC FT DOMAIN 75 88
 CC FT DNA_BIND 176 195 H-T-H MOTIF (BY SIMILARITY).
 CC FT SEQUENCE 212 AA; 22919 MW; 12A157F6F6B6BC3 CRC64;
 SQ
 Query Match 0.6%; Score 7; DB 1; Length 212;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 469 GAVRNAQ 475
 Db 188 GAVRNAQ 194
 RESULT 66
 ID R10_CAEEL STANDARD; PRT; 214 AA.
 AC 009533;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L10 (QM protein homolog).
 GN RPL-10 OR F10B5.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Poliderrinae; Caenorhabditis.
 CC NCBI_Taxid=62319;
 OK NCB1
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Simms M.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE L10E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC DR EMBL; 248334; CAA8308.1; -
 CC DR WormPep; P1085.1; CE01543.
 CC DR InterPro; IPR001197; Ribosomal_L10E.
 CC DR Pfam; PF00826; Ribosomal_L10E; 1.
 CC DR TIGRFAMs; TIGR00279; L10E; 1.
 CC DR PROSITE; PS01257; RIBOSOMAL_L10E; 1.
 CC KW Ribosomal protein.
 CC FT SEQUENCE 214 AA; 24749 MW; BC12D2B64BD31AAC CRC64;
 SQ
 Query Match 0.6%; Score 7; DB 1; Length 214;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1074 FRRAKFK 1080
 Db 152 FRRAKFK 158
 RESULT 67
 ID FGF3_CHICK STANDARD; PRT; 220 AA.
 AC P48801;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fibroblast growth factor-3 precursor (FGF-3) (HBGF-3).
 GN FGF3 OR FGF-3.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC OX NCB1_Taxid=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rhode Island red; TISSUE=Embryo;
 RX MEDLINE=95309122; PubMed=7789270;
 RA Mahmood R., Kiefer P., Guthrie S., Dickson C., Mason I.;
 RT "Multiple roles for FGF-3 during cranial neural development in the
 RT chicken.";
 RL Development 121:1399-1410 (1995).
 CC -1- FUNCTION: POTENT MITOGEN AND TRANSFORMING AGENT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 247555; CAA87635.1; -
 CC DR HSSP; P31371; I682.
 CC DR InterPro; IPR002209; HB/F_growthfact.
 CC DR InterPro; IPR002348; IIL_HBGF.
 CC DR Pfam; PF00167; FGF; 1.
 CC DR PRINTS; PR00262; IILHBGF.
 CC DR ProDom; PD000631; HB/F_growthfact; 1.
 CC DR SMART; SM00442; FGF; 1.
 CC DR PROSITE; PS00247; HBGF_FGF; 1.
 CC KW Growth factor; Mitogen; Signal; Glycoprotein.
 CC FT SIGNAL 1 19
 CC FT CHAIN 20 220 FIBROBLAST GROWTH FACTOR-3.
 CC FT CARBOHYD 66 66 N-LINKED (GLCNAC...) (POTENTIAL).
 CC FT SEQUENCE 220 AA; 25050 MW; B15D41D1551C5D5 CRC64;

FT DISULFID 198 213 BY SIMILARITY.
SQ SEQUENCE 223 AA; 24255 MW; 7D66D946DCADN00 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 223;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1135 LLLALLV 1141
|||||
Db 5 LLLALLV 11

RESULT 70
ID BRX1 HUMAN STANDARD; PRT; 225 AA.
AC 09HBL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein BarH-like 1.
GN BARX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND VARIANT ALA-19.
RC TISSUE=Craniocfacial;
RX MEDLINE=20453194; PubMed=10995576;
RA Gould D.B., Walter M.A.;
RT "Cloning, characterization, localization, and mutational screening of
the human BRX1 gene."
RL Genomics 68:336-342(2000).
CC -1- FUNCTION: TRANSCRIPTION FACTOR, WHICH MAY BE INVOLVED IN
CRANIOFACIAL DEVELOPMENT, IN ODONTOGENESIS AND IN STOMACH
ORGANOGENESIS. MAY HAVE A ROLE IN THE DIFFERENTIATION OF MOLARS
FROM INCISORS. BINDS TO A REGULATORY MODULE OF THE NCAM PROMOTER.
CC -1- SUBCELLULAR LOCATION: Nuclear (probable).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. EXPRESSED AT HIGHER LEVELS
IN TESTIS AND HEART. DETECTED IN CRANIOFACIAL TISSUE AND ADULT
IRIS, BUT NOT IN LYMPHOCYTES, FIBROBLASTS, CHOROIOID RETINA, RETINAL
PIGMENT EPITHELIUM, KIDNEY, OR FETAL LIVER.
CC -1- POLYMORPHISM: THE POLYMORPHISM IS NOT ASSOCIATED WITH AXENFELD-
REIER SYNDROME (ARS), IRIDOGONIOYSGENESIS SYNDROME (IGS) OR
RELATED OCULAR MALFORMATIONS.
CC -1- SIMILARITY: BELONGS TO THE BAR HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF213356; AAC23738.1; -;
DR HSSP; P14653; 1872.
DR Genew; HGNC:955; BARX1.
DR MIM; 603260; -;
DR InterPro; IPR000047; HTH_repressor.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PRO0024; HOMEBOX.
DR PRINTS; PRO0031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX 1; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KM Homeobox; Transcription; DNA-binding; Nuclear protein; Polymorphism.
FT DNA_BIND 113 172 HOMEBOX.
FT DOMAIN 15 24 POLY-ALA.
FT VARIANT 19 19 T->A.
/FTID=VAR_010927.

SQ SEQUENCE 225 AA; 24061 MW; A7907BB466F3393 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 225;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1019 LTPPGP 1025
|||||
Db 6 LTPPGP 12

RESULT 71
ID EMBP RAT STANDARD; PRT; 227 AA.
AC 063189;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Eosinophil granule major basic protein precursor (MBP).
GN PRG2 OR MBP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Bone marrow;
RX MEDLINE=96138543; PubMed=8547309; Misawa S., Isobe Y., Hayashi H.,
RA Nittoh T., Watanabe M., Okayama H.,
RA Mue S., Ohuchi K.;
RT "Cloning of cDNA for rat eosinophil major basic protein."
RL Biochim. Biophys. Acta 1264:261-264(1995).
CC -1- FUNCTION: CYTOTOXIN AND HELMINTHOTOXIN. MBP ALSO INDUCES
NONCYTOLYTIC HISTAMINE RELEASE FROM BASOPHILS. IT IS INVOLVED IN
ANTI-PARASITIC DEFENSE MECHANISMS AND IMMUNE HYPERSENSITIVITY
REACTIONS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
GRANULE (CRYSTALLOID CORE) (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D50568; BAA09129.1; -;
DR HSSP; P13727; 1H8U.
DR InterPro; IPR002352; Emaior_bas.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR PRINTS; PRO0770; EMBORBASICP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS00041; C-TYPE LECTIN 2; 1.
KM Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
Lectin.
FT SIGNAL 1 16 POTENTIAL.
FT PROPEP 17 110 ACIDIC (BY SIMILARITY).
FT CHAIN 111 227 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN.
FT DOMAIN 128 227 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 130 225 BY SIMILARITY.
FT DISULFID 202 217 BY SIMILARITY.
FT DOMAIN 59 68 POLY-GLU.
SQ SEQUENCE 227 AA; 25129 MW; 34F2B514090DE82 CRC64;
Query Match 0.6%; Score 7; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1135 LLLALLV 1141

```

Db          5 |||||
           5 LLLALLV 11

RESULT 72
NH40_CAEEL STANDARD; PRT; 231 AA.
ID NH40_CAEEL
AC Q22127;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear hormone receptor family member nhr-40.
GN NHR-40 OR 10366.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Tach A.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, U40940; AAA81704.2; -.
DR HSSP; P20393; 1A6Y.
DR WormPep; T03G6.2; CE28074.
DR InterPro; IPR000356; Hormone_rec_lig.
DR InterPro; IPR001628; ZnF_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; ZnF_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW zinc-finger.
FT DNA_BIND 15 77 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 15 35 C4-TYPE.
FT ZN_FING 51 72 C4-TYPE.
SQ SEQUENCE 231 AA; 26781 MW; EF68F2639494F2 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1157 EEEKREE 1163
DB 103 EEEKREE 109

RESULT 73
Y867_RICPR STANDARD; PRT; 231 AA.
ID Y867_RICPR
AC Q92C39;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Hypothetical protein RP867.
GN RP867.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiinae; Rickettsia.
OX NCBI_TaxId=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Madrid E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E.; Zomorodipour A.; Andersson J.O.;
RA Sierichitz-Ponten T.; Alsmark U.C.M.; Podowski R.M.; Naeslund A.K.;
RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria."
RL Nature 396:133-140(1998).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: STRONG, TO P.DENTRIFICANS HYPOTHETICAL 25.1 KDa
CC PROTEIN IN COBV 5'REGION (AC P29938).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL, AJ235273; CAA15291.1; -.
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
SQ SEQUENCE 231 AA; 26472 MW; 4B3FE686FA02343 CRC64;

Query Match 0.6%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 LGMSTLE 114
DB 209 LGMSTLE 215

RESULT 74
EMBL_CAVPO STANDARD; PRT; 233 AA.
ID EMBL_CAVPO
AC P22032;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Eosinophil granule major basic protein 1 precursor (MBP-1).
KW MBP.
GN Cavia porcellus (Guinea pig).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
OX NCBI_TaxId=10141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 115-162.
RC TISSUE=Eosinophil;
RX MEDLINE=91160746; PubMed=1705901;
RA Aoki I.; Shindoh Y.; Nishida T.; Nakai S.; Hong Y.-M.; Mio M.;
RA Saito T.; Tasaka K.;
RT "Sequencing and cloning of the cDNA of guinea pig eosinophil major
RT basic protein."
RL FEBS Lett. 279:330-334(1991).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=94092714; PubMed=8268206;
RA Hashimoto Y.; Nagaoaka I.; Yamashita T.;
RT "Purification of the antibacterial fragments of guinea-pig major
RT basic protein."
RL Biochim. Biophys. Acta 1203:236-242(1993).

```

```

CC -!- FUNCTION: MBP MAY PLAY SOME IMPORTANT ROLES IN THE ALLERGIC
CC REACTIONS AND INFLAMMATIONS. SINCE MBP IS CAPABLE OF RELEASING
CC HISTAMINE FROM MAST CELLS AND DAMAGING THE EPITHELIAL CELLS OF
CC BRONCHIAL TUBES. ANTIPARASITIC AND ANTI-BIOTIC.
CC -!- SUBCELLULAR LOCATION: MATRIX OF EOSINOPHIL'S LARGE SPECIFIC
CC GRANULE (CRYSTALLOID CORE).
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90251; BAA14291.1; -.
DR PIR; S13625; S13625.
DR HSSP; P13727; 1H8U.
DR InterPro; IPR002352; Emaior basic.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; Lectin_C; 1.
DR PRINTS; PR00770; EMAIORBASIC.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN 2; 1.
KW Eosinophil; Toxin; Signal; Cytotoxin; Immune response; Antibiotic;
KW Lectin; Multigene family.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 114 ACIDIC.
FT CHAIN 115 233 EOSINOPHIL GRANULE MAJOR BASIC PROTEIN 1.
FT DOMAIN 132 233 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 134 231 BY SIMILARITY.
FT DISULFID 208 223 BY SIMILARITY.
SQ SEQUENCE 233 AA; 26268 MW; C8D5E96D927C56C8 CRC64;

```

```

Query Match 0.6%; Score 7; DB 1; Length 233;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1134 LLLALL 1140
DB 4 LLLALL 10

```

```

RESULT 75
COMB_CLOAB
ID COMB_CLOAB STANDARD; PRT; 235 AA.
AC Q97E82;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable 2-phosphosulfolactate phosphatase (EC 3.1.3.-).
GN COMB OR CAC3233.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1488;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=2159325; PubMed=146286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabatne F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
CC -!- CATALYTIC ACTIVITY: 2-phosphosulfolactate = sulfolactate +
CC phosphate.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- SIMILARITY: BELONGS TO THE COMB FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE007819; AAK81168.1; -.
KW Hydrolase; Magnesium; Complete proteome.
SQ SEQUENCE 235 AA; 26144 MW; ABBFB7DF07161EBD CRC64;

```

```

Query Match 0.6%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1125 ILIGSVL 1131
DB 118 ILIGSVL 124

```

```

Search completed: July 16, 2003, 08:32:39
Job time : 19 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:30:30 ; Search time 29 Seconds
(without alignments)
3868.582 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167
Sequence: 1 MELPVTHTLPLVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	1.1	272	2 A55348	integrin alpha-1 -
2	13	1.1	1151	2 A45226	integrin alpha-1 c
3	13	1.1	1180	2 A35854	integrin alpha-1 c
4	11	0.9	191	2 I47230	VLA-2 protein - pi
5	11	0.9	687	2 C71849	carbon starvation
6	11	0.9	687	2 H64665	carbon starvation
7	11	0.9	1170	2 I45914	integrin alpha 2 s
8	11	0.9	1178	2 S44142	VLA-2 protein homo
9	11	0.9	1181	2 A33998	integrin alpha-2 c
10	10	0.9	1179	2 A53213	integrin alpha-E c
11	9	0.8	84	2 B95310	hypothetical prote
12	9	0.8	104	2 PH0141	hypothetical prote
13	9	0.8	236	2 T12766	hypothetical prote
14	9	0.8	1163	2 I56126	probable lipoprote
15	9	0.8	1170	2 S03308	lymphocyte functio
16	8	0.7	42	2 A60822	cell surface glyco
17	8	0.7	20	2 A34259	cytochrome P450 PB
18	8	0.7	76	2 A43537	cytochrome P450mc4
19	8	0.7	76	2 I53107	heat-stable antige
20	8	0.7	80	2 A48966	CD24 precursor - r
21	8	0.7	103	2 A60856	B cell surface ant
22	8	0.7	127	2 B75301	conserved hypothet
23	8	0.7	136	2 G82058	hypothetical prote
24	8	0.7	141	2 S15785	ribosomal protein
25	8	0.7	169	2 AB1010	heat-stable antige
26	8	0.7	185	2 H96977	formylmethionine d
27	8	0.7	234	1 S15102	probable membrane
28	8	0.7	240	2 E72629	eosinophil major b
29	8	0.7	251	2 A81345	hypothetical prote
					histidine-binding

30	8	0.7	254	2 B55904	C 3.4.25.1 proteas
31	8	0.7	259	2 T37985	proteasome compone
32	8	0.7	264	2 S22090	catechol O-methyl
33	8	0.7	281	2 S18541	hypothetical prote
34	8	0.7	298	2 F71556	probable had hydro
35	8	0.7	309	2 C65217	allose kinase (EC
36	8	0.7	360	2 A85016	hypothetical prote
37	8	0.7	367	2 S19172	cytochrome P450 2B
38	8	0.7	383	2 T31771	hypothetical prote
39	8	0.7	415	2 G83568	probable permease
40	8	0.7	433	2 AC2275	hypothetical prote
41	8	0.7	444	2 T01721	hypothetical prote
42	8	0.7	491	1 OARTPB	cytochrome P450 2B
43	8	0.7	491	1 OARTP2	cytochrome P450 2B
44	8	0.7	500	2 B31047	testosterone 16alp
45	8	0.7	572	2 I39311	Kruppel-type zinc
46	8	0.7	604	2 I36917	glycoprotein Iib -
47	8	0.7	607	2 S60658	legum - Gnaetum g
48	8	0.7	614	2 A69845	Na+/H+ antiporter
49	8	0.7	685	2 AC0527	ferrichrome transp
50	8	0.7	767	2 G36661	hypothetical prote
51	8	0.7	849	2 E83349	probable ClpA/B-Y
52	8	0.7	942	1 J01674	protein kinase TMK
53	8	0.7	959	1 B60017	outer capsid prote
54	8	0.7	969	2 T15446	hypothetical prote
55	8	0.7	1037	2 A60163	glycoprotein Iib -
56	8	0.7	1039	2 A34269	integrin alpha-2b
57	8	0.7	1146	2 S40311	integrin - fruit f
58	8	0.7	1153	2 RWHU1B	cell surface glyco
59	8	0.7	1163	1 RWHU1C	cell surface glyco
60	8	0.7	1500	2 G84922	hypothetical prote
61	8	0.7	2109	2 I38414	transcription fact
62	8	0.7	2222	2 T13924	sdh protein - frui
63	8	0.7	3176	1 CGHU3A	collagen alpha 3(V
64	7	0.6	47	2 D81666	hypothetical prote
65	7	0.6	73	2 C65069	hypothetical prote
66	7	0.6	74	2 T10456	dephosphatase B3 pr
67	7	0.6	76	2 I51527	integrin alpha 5 s
68	7	0.6	77	2 I51529	integrin alpha V-1
69	7	0.6	84	2 D70640	hypothetical prote
70	7	0.6	85	2 G82776	ACP xP0672 import
71	7	0.6	86	2 B10993	hypothetical prote
72	7	0.6	86	2 G85938	hypothetical prote
73	7	0.6	89	2 I49515	B144 protein B - m
74	7	0.6	94	2 T03285	yeast-specific pr
75	7	0.6	97	1 A64146	hypothetical prote
76	7	0.6	100	2 F72587	insulin precursor
77	7	0.6	107	1 IPEH	insulin alpha-7C
78	7	0.6	108	2 S40149	hypothetical prote
79	7	0.6	109	2 H72713	protein hbaa precu
80	7	0.6	110	2 S30268	protein hbaa precu
81	7	0.6	110	2 F91177	protein hbaa precu
82	7	0.6	110	2 G86023	protein hbaa precu
83	7	0.6	111	2 T36386	probable Lsr2-like
84	7	0.6	115	2 D75434	T-cell receptor be
85	7	0.6	119	2 B32578	hypothetical prote
86	7	0.6	127	2 S76755	hypothetical prote
87	7	0.6	129	2 F72488	hypothetical prote
88	7	0.6	130	2 D83305	hypothetical prote
89	7	0.6	132	2 B81215	conserved hypothet
90	7	0.6	133	2 H75386	hypothetical prote
91	7	0.6	137	2 P00502	hypothetical prote
92	7	0.6	139	2 A38612	insulin-like growt
93	7	0.6	142	1 S59391	hypothetical prote
94	7	0.6	142	2 UT0573	retinoic acid-indu
95	7	0.6	142	2 G87647	hypothetical prote
96	7	0.6	143	2 T43180	hypothetical prote
97	7	0.6	143	2 T40426	hypothetical prote
98	7	0.6	144	2 A44585	globin - sea hare
99	7	0.6	145	2 B72724	hypothetical prote
100	7	0.6	148	2 D75090	hypothetical prote

ALIGNMENTS

RESULT 1

A55348
Integrin alpha-1 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
C:Accession: A55348
R:Kern, A.; Brissewitz, R.; Bank, I.; Marcantonio, E.E.
J. Biol. Chem. 269, 22811-22816, 1994
A:Title: The role of the I domain in ligand binding of the human integrin alpha-1beta-1.
A:Reference number: A55348; PMID:94357930; PMID:7521332
A:Accession: A55348
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-272 <KER>
A:Cross-references: GB:U10114
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
F:55-230/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match
Best Local Similarity 100.0%; Score 13; DB 2; Length 272;
Pred. No. 0.00016;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
DB 59 VIVLDGNSIYPM 71

RESULT 2

A45226
Integrin alpha-1 chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2000
C:Accession: A45226
R:Brissewitz, R.; Epstein, M.R.; Marcantonio, E.E.
J. Biol. Chem. 268, 2989-2996, 1993
A:Title: Expression of native and truncated forms of the human integrin alpha 1 subunit.
A:Reference number: A45226; PMID:93155124; PMID:8428973
A:Accession: A45226
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1151 <BRI>
A:Experimental source: hepatoblastoma cell line HepG2
A:Note: sequence extracted from NCBI backbone (NCBIP:124326)
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
F:142-317/Domain: von Willebrand factor type A repeat homology <VWA1>

Query Match
Best Local Similarity 100.0%; Score 13; DB 2; Length 1151;
Pred. No. 0.00052;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
DB 146 VIVLDGNSIYPM 158

RESULT 3

A35854
Integrin alpha-1 chain precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 23-Oct-1990 #sequence_revision 13-Sep-1991 #text_change 20-Sep-1999
C:Accession: A35854; S11243
R:Ignatius, M.J.; Large, T.H.; Houde, M.; Tawil, J.W.; Barton, A.; Esch, F.; Carbonetto, J. Cell Biol. 111, 709-720, 1990
A:Title: Molecular cloning of the rat integrin alpha-1 subunit: a receptor for laminin A
A:Reference number: A35854; PMID:90338125; PMID:2380249
A:Accession: A35854
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1180 <IGN>
A:Cross-references: GB:X52140; NID:956493; PIDN:CAA36384.1; PID:956494

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
C:Keywords: cell adhesion; cytoskeleton; transmembrane protein
F:170-345/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match
Best Local Similarity 100.0%; Score 13; DB 2; Length 1180;
Pred. No. 0.00053;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
DB 174 VIVLDGNSIYPM 186

RESULT 4

I47230
VLA-2 protein - pig (fragment)
A:Alternate names: glycoprotein Ia/IIa
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Sep-1999
C:Accession: I47230; S21518
R:Bahou, W.F.; Potter, C.L.; Mirza, H.
Blood 84, 3734-3741, 1994
A:Title: The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific recognition
A:Reference number: I47230; PMID:95036279; PMID:7949129
A:Accession: I47230
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-191 <BAH>
A:Cross-references: EMBL:Z12137; NID:92158; PIDN:CAA78125.1; PID:92159
C:Superfamily: unassigned collagens
C:Keywords: glycoprotein

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 191;
Pred. No. 0.013;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
DB 92 VVVTGESHHDG 102

RESULT 5

C71849
Carbon starvation protein A homolog jhp1095 [similarity] - Helicobacter pylori (strain J
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C:Accession: C71849
R:Alm, R.A.; Lang, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mille, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; PMID:99120557; PMID:9923682
A:Accession: C71849
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-687 <ARN>
A:Cross-references: GB:AE001537; GB:AE001439; NID:94155687; PIDN:AAD06677.1; PID:9415569
A:Experimental source: strain J99
C:Superfamily: carbon starvation protein

Query Match
Best Local Similarity 100.0%; Score 11; DB 2; Length 687;
Pred. No. 0.037;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMIILGSLVLCG 1133
DB 118 LMIILGSLVLCG 128

RESULT 6

1123 LMIILGSLVLCG 1133
118 LMIILGSLVLCG 128

H64665
Carbon starvation protein A homolog HP1168 [similarity] - Helicobacter pylori (strain 26
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C:Accession: H64665
R:Tombl, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodok, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Kark, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: H64665
A:Molecule type: DNA
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Residues: 1-687 <TOM>
A:Cross-references: GB:AE000622; GB:AE000511; NID:g2314317; PIDN:AAD08212.1; PID:g231432
C:Superfamily: Carbon starvation protein

Query Match 0.9%; Score 11; DB 2; Length 687;
Best Local Similarity 100.0%; Pred. No. 0.037;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LMIIGSVVGG 1133
|||||
Db 118 LMIIGSVVGG 128

RESULT 7

I15914
Integrin alpha 2 subunit - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 20-Sep-1999
C:Accession: I45914
R:Kamata, T.; Puzon, W.; Takada, Y.
J. Biol. Chem. 269, 9659-9663, 1994
A:Title: Identification of putative ligand binding sites within the I-domain of integrin
A:Reference number: A54402; MUID:94193647; PMID:7511592
A:Accession: I45914
A:Molecule type: mRNA
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-1170 <KAM>
A:Cross-references: GB:I25886; NID:9439695; PIDN:AAB59235.1; PID:9439696
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
F:161-336/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 0.9%; Score 11; DB 2; Length 1170;
Best Local Similarity 100.0%; Pred. No. 0.058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
|||||
Db 268 VVVTGESHHDG 278

RESULT 8

VLA-2 protein homolog - mouse
S44142
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Sep-1999
C:Accession: S44142
R:Edelman, J.M.; Chan, B.M.; Uneyal, S.; Onodera, H.; Wang, D.Z.; Damjanovich, L.; Latze
submitted to the EMBL Data Library, January 1994
A:Description: The mouse VLA-2 homologue supports collagen and laminin adhesion but not
A:Reference number: S44142
A:Accession: S44142
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1178 <EDE>
A:Cross-references: EMBL:229987; NID:9473098; PIDN:CAA82877.1; PID:9473099
C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
F:163-344/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match 0.9%; Score 11; DB 2; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
|||||
Db 276 VVVTGESHHDG 286

RESULT 9

A33998
Integrin alpha-2 chain precursor - human
N:Alternate names: CD49b; platelet glycoprotein GPIa; VLA-2/collagen receptor alpha-2 ch.
C:Species: Homo sapiens (man)
C:Date: 30-Mar-1990 #sequence_revision 18-Sep-1992 #text_change 21-Jul-2000
C:Accession: A33998; B56793; A53117
R:Takada, Y.; Hemler, M.E.
J. Cell. Biol. 109, 397-407, 1990
A:Title: The primary structure of the VLA-2/collagen receptor alpha(2)-subunit (platelet
A:Reference number: A33998; MUID:89308879; PMID:2545729
A:Accession: A33998
A:Molecule type: mRNA
A:Residues: 1-1181 <TK>
A:Cross-references: GB:X17033; NID:933906; PIDN:CAA34894.1; PID:933907
A:Note: the authors translated the codon GAT for residue 802 as Gln, GTC for residue 803
R:Catmell, B.; Parentier, S.; Leung, L.L.; McGregor, J.L.
Biochem. J. 279, 419-425, 1991
A:Title: Separation of important new platelet glycoproteins (GPIa, GPIC, GPIC*, GP1a an
A:Reference number: B56793; MUID:92061944; PMID:1955640
A:Accession: B56793
A:Molecule type: protein
A:Residues: 30-43 <CAT>
A:Experimental source: platelet
R:Zutter, M.M.; Santoro, S.A.; Painter, A.S.; Tsung, Y.L.; Gafford, A.
J. Biol. Chem. 269, 463-469, 1994
A:Title: The human alpha-2 integrin gene promoter. Identification of positive and negati
A:Reference number: A53117; MUID:94103255; PMID:8276836
A:Accession: A53117
A:Molecule type: DNA
A:Residues: 1-16, 'V', 18-21 <ZUT>
A:Cross-references: GB:I24121; NID:9400342; PIDN:AA16619.2; PID:94583535
A:Note: authors translated the codon GTA for residue 17 as Leu
C:Genetics:
A:Gene: GDB:ITGA2; CD49B
A:Cross-references: GDB:128031; OMIM:192974
A:Map position: 5q11.1-5q11.2

C:Superfamily: unassigned collagens; von Willebrand factor type A repeat homology
C:Keywords: cell adhesion; glycoprotein; heterodimer; transmembrane protein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-1133/Domain: extracellular #status predicted <EXT>
F:172-347/Domain: von Willebrand factor type A repeat homology <VWA2>
F:1134-1154/Domain: transmembrane #status predicted <TM>
F:1155-1181/Domain: intracellular #status predicted <CT>
F:105,112,343,432,460,475,699,1057,1074,1081/Binding site: carbohydrate (Asn) (covalent)

Query Match 0.9%; Score 11; DB 2; Length 1181;
Best Local Similarity 100.0%; Pred. No. 0.058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGESHHDG 282
|||||
Db 279 VVVTGESHHDG 289

RESULT 10

A53213
Integrin alpha-E chain - human
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: A53213
R:Shaw, S.K.; Ceppek, K.L.; Murphy, E.A.; Russell, G.J.; Brenner, M.B.; Parker, C.M.
J. Biol. Chem. 269, 6016-6023, 1994
A:Title: Molecular cloning of the human mucosal lymphocyte integrin alpha(E) subunit. Un

A:Reference number: A53213; MUID:94164962; PMID:8119947
 A:Accession: A53213
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1179 <SHA>
 A:Cross-references: GB:U5851; NID:9457244; PID:9457245
 C:Genetics:
 A:Gene: GDB:ITGAE
 A:Cross-references: GDB:330801
 A:Map position: 17p13
 C:Superfamily: cell surface glycoprotein CD11b, von Willebrand factor type A repeat homology
 F:199-371/Domain: von Willebrand factor type A repeat homology <WMA3>

Query Match 0.9%; Score 10; DB 2; Length 1179;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYGSELCP 492
 |||||
 Db 511 GSYGSELCP 520

RESULT 11
 B95310
 hypotetical protein SMA0726 [imported] - Sinorhizobium meliloti (strain 1021) magaplastm
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: B95310
 R:Karnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bower, J.; Katman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: B95310
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-84 <KUN>
 A:Cross-references: GB:AE006469; PIDN:AAK65044.1; PID:914523476; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplastid pSymA
 R:Galbert, F.; Finn, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chai, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Leilaure, hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMA0726
 A:Genome: plasmid

Query Match 0.8%; Score 9; DB 2; Length 84;
 Best Local Similarity 100.0%; Pred. No. 0.72;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 RLSPRLRL 729
 |||||
 Db 58 RLSPRLRL 66

RESULT 12
 PH0141
 integrin alpha chain - chicken (fragments)
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Apr-1999
 C:Accession: PH0141
 R:Syfrig, J.; Mann, K.; Paulsson, M.
 Exp. Cell Res. 194, 165-173, 1991
 A:Title: An abundant chick gizzard integrin is the avian alpha1-beta1 integrin heterodimer
 A:Reference number: PH0141; MUID:91224161; PMID:1651093
 A:Accession: PH0141
 A:Molecule type: protein

A:Residues: 1-104 <SYF>
 A:Experimental source: gizzard
 C:Comment: Integrins are a family of cellular adhesion receptors that mediate interaction
 C:Superfamily: unassigned collagens
 C:Keywords: cell adhesion

Query Match 0.8%; Score 9; DB 2; Length 104;
 Best Local Similarity 100.0%; Pred. No. 0.86;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEQIGSYRG 487
 |||||
 Db 43 GEQIGSYRG 51

RESULT 13
 T12766
 probable lipoprotein yoxB - Bacillus subtilis phage SPBc2
 C:Species: Bacillus subtilis phage SPBc2
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
 C:Accession: T12766; F69907
 R:Izazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.
 submitted to the EMBL Data Library, August 1997
 A:Description: The complete nucleotide sequence of the Bacillus subtilis SPBc2 prophage
 A:Reference number: 217583
 A:Accession: T12766
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-236 <LAZ>
 A:Cross-references: EMBL:AF020713; NID:93025478; PID:93025480; PIDN:AAI12975.1
 R:Kunat, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bettec, C.; Bron, S.; Brunnelle, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, S.; Galizzi, A.; Gallier, lech, J.; Harwood, C.R.; Hentut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, A.; Oudga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Scanlon, A.; Authors: Schelke, S.; Schreier, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seror, akuch, M.; Tanakoshi, A.; Tanaka, T.; Terpetra, F.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Whalers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumelein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: F69907
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-236 <KUN>
 A:Cross-references: GB:Z59115; GB:AL009126; NID:92634478; PIDN:CAB14083.1; PID:ell183612;
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yoxB

Query Match 0.8%; Score 9; DB 2; Length 236;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164
 |||||
 Db 209 PEEKREK 217

RESULT 14
 I56126
 lymphocyte function-associated molecule-1-alpha - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Aug-1999
 C:Accession: I56126
 R:Kautmann, Y.; Tseng, E.; Springer, T.A.
 J. Immunol. 147, 369-374, 1991
 A:Title: Cloning of the murine lymphocyte function-associated molecule-1 alpha-subunit a

```

1
A:Reference number: 156126; MUID:91268576; PMID:2051027
A:Accession: 156126
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1163 <RES>
A:Cross-references: GB:M60778; NID:9198785; PIDN:AAA9426.1; PID:9198786
C:Gene: LFA-1
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology
F:151-315/Domain: von Willebrand factor type A repeat homology <VWA1>

Query Match          0.8%; Score 9; DB 2; Length 1163;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574
DB 536 DVAVGAPLE 544

RESULT 15
S03308
cell surface glycoprotein CD11a precursor - human
M:Alternate names: leukocyte adhesion glycoprotein LFA-1 alpha chain; leukocyte function
C:Species: Homo sapiens (man)
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 20-Aug-1999
C:Accession: S03308; A47458; A47565; A48759; S36044
R:Larsen, R.S.; Corbi, A.L.; Berman, L.; Springer, T.
J:Cell Biol. 108, 703-712, 1989
A:Title: Primary structure of the leukocyte function-associated molecule-1 alpha subunit
A:Reference number: S03308; MUID:89139587; PMID:2537322
A:Accession: S03308
A:Molecule type: mRNA
A:Residues: 1-1170 <LAR>
A:Cross-references: EMBL:Y00796; NID:g31421; PIDN:CAA68747.1; PID:g31422
A>Note: part of this sequence was confirmed by protein sequencing
R:Corrwell, R.D.; Gollahan, K.A.; Hackett, D.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 4221-4225, 1993
A:Title: Description of the leukocyte function-associated antigen 1 (LFA-1 or CD11a) pro
A:Reference number: A47458; MUID:93248261; PMID:8097887
A:Accession: A47458
A:Molecule type: DNA
A:Residues: 1-20 <COR>
A>Note: sequence extracted from NCBI backbone (NCBIN:130862; NCBI:130863)
R:Shelley, C.S.; Parthazad, O.C.; Arnaout, M.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 5364-5368, 1993
A:Title: Identification of cell-specific and developmentally regulated nuclear factors b
A:Reference number: A47565; MUID:93281759; PMID:8099450
A:Accession: A47565
A:Molecule type: DNA
A:Residues: 1-20 <SHE>
R:Nueda, A.; Lopez-Cabrera, M.; Vaz, A.; Corbi, A.L.
J. Biol. Chem. 268, 19305-19311, 1993
A:Title: Characterization of the CD11a (alphaL, LFA-1alpha) integrin gene promoter.
A:Reference number: A48759; MUID:93374910; PMID:8103515
A:Accession: A48759
A:Molecule type: DNA
A:Residues: 1-20 <NUE>
A:Cross-references: EMBL:Z22804; NID:g311405; PIDN:CAA80461.1; PID:g311406
C:Gene: GDB:ITGAL; CD11A
A:Cross-references: GDB:119757; OMIM:153370
A:Map position: 16p11.2-16p11.2
C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homology
C:Keywords: cell adhesion; cytoskeleton; glycoprotein; heterodimer; surface antigen; tra
F:1-25/Domain: signal sequence status predicted <SIG>
F:26-1170/Product: leukocyte adhesion glycoprotein LFA-1 alpha chain #status predicted <
F:154-317/Domain: von Willebrand factor type A repeat homology <VWA2>

Query Match          0.8%; Score 9; DB 2; Length 1170;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 566 DVAVGAPLE 574
DB 538 DVAVGAPLE 546

RESULT 16
A60822
cytochrome P450 PB-3a - rat (fragment)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jun-1999
C:Accession: A60822; 155191
R:Ameliazad, Z.; Nardonne, J.F.; Wolf, C.R.; Robertson, L.W.; Oesch, F.
Biochem. Pharmacol. 37, 3245-3249, 1988
A:Title: Effect of nutritional imbalances on cytochrome P-450 isozymes in rat liver.
A:Reference number: A60822; MUID:88293549; PMID:3041969
A:Accession: A60822
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-20 <AME>
R:Hashimoto, T.; Matsumoto, T.; Nishizawa, M.; Kawabata, S.
J. Biochem. 103, 487-492, 1988
A:Title: A mutant rat strain deficient in induction of a phenobarbital-inducible form o
A:Reference number: 155191; MUID:88273074; PMID:2839467
A:Accession: 155191
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-20 <RES>
A:Cross-references: GB:D00250; NID:g220727; PIDN:BA400181.1; PID:g220728
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: electron transfer; heme; monooxygenase; oxidoreductase; transmembrane protei

Query Match          0.7%; Score 8; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLLV 1141
DB 6 LLLALLLV 13

RESULT 17
A34259
cytochrome P450mt4, phenobarbital-inducible, mitochondrial, hepatic - rat (fragment)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 05-Mar-1999
C:Accession: A34259
R:Shayiq, R.M.; Avadhani, N.G.
Biochemistry 29, 866-873, 1990
A:Title: A phenobarbital-inducible hepatic mitochondrial cytochrome P-450 immunochemical
A:Reference number: A34259; MUID:90254127; PMID:2340279
A:Accession: A34259
A:Molecule type: protein
A:Residues: 1-42 <SHA>
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: electron transfer; heme; liver; mitochondrion; monooxygenase; oxidoreductase

Query Match          0.7%; Score 8; DB 2; Length 42;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLLV 1141
DB 6 LLLALLLV 13

RESULT 18
A43537
heat-stable antigen M1/69-J11d precursor - mouse
M:Alternate names: CD24 protein; necladrin
C:Species: Mus musculus (house mouse)

```

C>Date: 06-Nov-1992 #sequence, revision 06-Nov-1992 #text, change 31-Jan-2000
C:Accession: A43537; 148287; S15784; S15783; S43709; S32240; S33129
R:Kay, R.; Takei, F.; Humphries, R.K.
J. Immunol. 145, 1952-1959, 1990
A:Title: Expression cloning of a cDNA encoding M1/69-J11d heat-stable antigens.
A:Reference number: A43537; MUID:90361906; PMID:2118158
A:Accession: A43537
A:Molecule type: mRNA
A:Residues: 1-76 <KX>
A:Cross-references: GB:M58661; NID:g198985; PIDN:AA39481.1; PID:g198986
R:Wenger, R.H.; Rochelle, J.M.; Seidin, M.F.; Kohler, G.; Nielsen, P.J.
J. Biol. Chem. 268, 23345-23352, 1993
A:Title: The heat stable antigen (mouse CD24) gene is differentially regulated but has a
A:Reference number: A48876; MUID:94043137; PMID:8226859
A:Accession: 148287
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-76 <RES>
A:Cross-references: EMBL:X72910; NID:g296466; PIDN:CAA51415.1; PID:g296467
R:Wenger, R.H.; Ayane, M.; Bose, R.; Koehler, G.; Nielsen, P.J.
Eur. J. Immunol. 21, 1039-1046, 1991
A:Title: The genes for a mouse hematopoietic differentiation marker called the heat-stab
A:Reference number: S15783; MUID:91209380; PMID:2019286
A:Accession: S15784
A:Molecule type: DNA
A:Residues: 1-76 <WE3>
A:Cross-references: NID:g51439; PIDN:CAA39841.1; PID:g51440
A:Accession: S15783
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 32-76 <WE2>
A:Cross-references: EMBL:X53825
R:Nielsen, P.J.
submitted to the EMBL Data Library, July 1990
A:Reference number: S19111
A:Accession: S43709
A:Molecule type: mRNA
A:Residues: 1-76 <NIE>
A:Cross-references: EMBL:X53825; NID:g55441; PIDN:CAA37822.1; PID:g55442
C:Genetics:
A:Gene: CD24
A:Map position: 10
A:Introns: 23/3
C:Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphati
F:1-26/Domain: signal sequence #status predicted <Sig>
F:27-56/Product: heat-stable antigen M1/69-J11d #status predicted <MAT>
F:57-76/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:56/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)
Query Match 0.7%; Score 8; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1133 GULLALL 1140
DB 12 GULLALL 19
RESULT 19
153107
CD24 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence, revision 02-Aug-1996 #text, change 05-Nov-1999
C:Accession: I53107; S25146
R:Shirasawa, T.; Akashi, T.; Sakamoto, K.; Takahashi, H.; Maruyama, N.; Hirokawa, K.
Dev. Dyn. 198, 1-13, 1993
A:Title: Gene expression of CD24 core peptide molecule in developing brain and developin
A:Reference number: I53107; MUID:94122434; PMID:8292828
A:Accession: I53107
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-76 <RES>
A:Cross-references: EMBL:211663; NID:g55901; PIDN:CAA7731.1; PID:g55902

C:Keywords: phosphatidylinositol linkage
Query Match 0.7%; Score 8; DB 2; Length 76;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1133 GULLALL 1140
DB 12 GULLALL 19
RESULT 20
A48996
B cell surface antigen CD24 precursor - human
N:Alternate names: cluster-w4 antigen; signal transducer CD24
C:Species: Homo sapiens (man)
C>Date: 19-Dec-1993 #sequence, revision 02-Aug-1996 #text, change 21-Jul-2000
C:Accession: I56114; A48996; I54201; I72676
R:Kay, R.; Rosten, P.M.; Humphries, R.K.
J. Immunol. 147, 1412-1416, 1991
A:Title: CD24, a signal transducer modulating B cell activation responses, is a very sho
A:Reference number: I56114; MUID:91332458; PMID:1831224
A:Accession: I56114
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-80 <RES>
A:Cross-references: GB:I33930; NID:g500848; PID:g500849; GB:M58664; NID:g180167; PID:g18
R:Jackson, D.; Walpel, R.; Weber, E.; Bell, J.; Stahel, R.A.
Cancer Res. 52, 5264-5270, 1992
A:Title: CD24, a signal-transducing molecule expressed on human B cells, is a major surf
A:Reference number: A48996; MUID:93007871; PMID:1327504
A:Accession: A48996
A:Molecule type: mRNA
A:Residues: 1-56, 'V', 58-80 <JAC>
A:Cross-references: GB:X63397; GB:S44888; NID:g396167; PIDN:CAA49195.1; PID:g396168
A:Experimental source: small cell lung carcinoma line SW2
A:Note: sequence extracted from NCBI Backbone (NCBIN:114635, NCBI:P:114636)
R:Hough, M.R.; Rosten, P.M.; Sexton, T.L.; Kay, R.; Humphries, R.K.
Genomics 22, 154-161, 1994
A:Title: Mapping of CD24 and homologous sequences to multiple chromosomal loci.
A:Reference number: I54201; MUID:95048364; PMID:7959762
A:Accession: I54201
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 5-11, 'W', 13-43, 'T', 45-80 <RE2>
A:Cross-references: GB:S75311; NID:g833885; PIDN:AA014170.1; PID:g4261870
C:Genetics:
A:Gene: GDB:CD24
A:Cross-references: GDB:383795; OMIM:600074
A:Map position: 6q21-6q21
C:Keywords: B-cell; blocked carboxyl end; glycoprotein; lipoprotein; phosphatidylinosito
F:1-22/Domain: signal sequence #status predicted <Sig>
F:22-67/Product: B cell surface antigen CD24 #status predicted <MAT>
F:68-80/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:67/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)
Query Match 0.7%; Score 8; DB 2; Length 80;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1133 GULLALL 1140
DB 12 GULLALL 19
RESULT 21
AF0856
conserved hypothetical protein STY3056 [imported] - Salmonella enterica subsp. enterica
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence, revision 09-Nov-2001 #text, change 27-Nov-2001
C:Accession: AF0856

R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, th, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0856
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <PAR>
A:Cross-References: GB:AL513382; PIDN:CAD06037.1; PID:g16504004; GSPDB:GN00176
C:Genetics:
A:Gene: STY3056
C:Superfamily: hypothetical protein HI0673

Query Match 0.7%; Score 8; DB 2; Length 103;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
|||||
Db 6 LLLALLLV 13

RESULT 22
B75301
hypothetical protein - *Deinococcus radiodurans* (strain R1)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
C:Accession: B75301
R.White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.U.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zaluski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; PMID:20036896; PMID:10567266
A:Accession: B75301
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <WHI>
A:Cross-References: GB:AE002054; GB:AE000513; NID:g6460010; PIDN:AAF11763.1; PID:g646001
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR2210
A:Map position: 1
C:Superfamily: *Deinococcus radiodurans* hypothetical protein DR2210

Query Match 0.7%; Score 8; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
|||||
Db 7 GLLALL 14

RESULT 23
G82058
ribosomal protein L16 VC2589 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)
C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: G82058
R.Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.U.;
chardson, D.; Esmolaeva, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F
I.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; PMID:20406833; PMID:10952301
A:Accession: G82058
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-136 <HEI>

A:Cross-References: GB:AE004326; GB:AE003852; NID:g9657161; PIDN:AAF95730.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2589
A:Map position: 1
C:Superfamily: *Escherichia coli* ribosomal protein L16

Query Match 0.7%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1055 LAKGTVEVS 1062
|||||
Db 20 LAKGTVEVS 27

RESULT 24
S15785
heat-stable antigen-related hypothetical protein HSA-C - mouse
C:Species: *Mus musculus* (house mouse)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S15785
R.Wenger, R.H.; Ayane, M.; Bose, R.; Koehler, G.; Nielsen, P.J.
Eur. J. Immunol. 21, 1039-1046, 1991
A:Title: The genes for a mouse hematopoietic differentiation marker called the heat-stab
A:Reference number: S15783; PMID:91209380; PMID:2019286
A:Accession: S15785
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <WEN>
A:Cross-References: EMBL:X56486; NID:g51441; PIDN:CAA39843.1; PID:g51442
A:Note: the authors translated the codon TTC for residue 87 as Ser and AGA for residue 1
A:Note: the authors did not translate the codon for residue 101
C:Genetics:
A:Introns: #status absent
C:Keywords: glycoprotein; membrane protein; phosphatidylinositol linkage

Query Match 0.7%; Score 8; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
|||||
Db 12 GLLALL 19

RESULT 25
AB1010
formylmethionine deformylase (EC 3.5.1.31) - *Salmonella enterica* subsp. *enterica* serovar
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB1010
R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AB1010
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <PAR>
A:Cross-References: GB:AL513382; PIDN:CAD09179.1; PID:g16505183; GSPDB:GN00176
C:Genetics:
A:Gene: fms
C:Superfamily: polypeptide deformylase
C:Keywords: hydrolase

Query Match 0.7%; Score 8; DB 2; Length 169;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198
 |||||
 DB 138 LVGKLFID 145

RESULT 26

H96977
 Probable membrane protein (imported) - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: H96977
 R:Willing, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium ClC
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H96977
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-185 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK78611.1; PID:G15023507; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 A:Genetics: CAC0634

Query Match 0.7%; Score 8; DB 2; Length 185;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 636 AAILSSR 643
 |||||
 DB 54 AAILSSR 61

RESULT 27
 S15102
 eosinophil major basic protein 2 precursor - guinea pig
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
 C:Accession: S15102; S18501
 R:AOX, I.; Shindoh, Y.; Nishida, T.; Nakai, S.; Hong, Y.M.; Mio, M.; Saito, T.; Tasaka, FBS Lett. 282, 56-60, 1991
 A:Title: Comparison of the amino acid and nucleotide sequences between human and two guinea pig eosinophil major basic proteins
 A:Reference number: S15102; MUID:91224343; PMID:2026266
 A:Accession: S15102
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-234 <AOX>
 A:Cross-references: DBJ:DD0817; NID:G3135095; PIDN:BAA00697.1; PID:G220293
 A:Accession: S18501
 A:Molecule type: protein
 A:Residues: 116, 'X', 118-134, 'X', 136-137, 'X', 139-145, 161-176, 181-200 <AOX>
 C:Superfamily: eosinophil major basic protein precursor; C-type lectin homology
 C:Keywords: antibiologic; chondroitin sulfate proteoglycan; cytolysin; eosinophil; glycoprotein
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:117-115/Domain: signal sequence #status predicted <PRO>
 F:116-234/Product: eosinophil major basic protein 2 #status experimental <MAT>
 F:117-232/Domain: C-type lectin homology <LCH>
 F:24,25/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:69/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 F:335-232,209-224/Disulfide bonds: #status predicted

Query Match 0.7%; Score 8; DB 1; Length 234;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 |||||
 DB 4 LLLALLV 11

RESULT 28

E72629
 hypothetical protein APE1493 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
 C:Accession: E72629
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: E72629
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-240 <KAW>
 A:Cross-references: DBJ:AP000061; NID:G5104821; PIDN:BAA80491.1; PID:dl044277; PID:G510
 A:Experimental source: strain K1
 A:Genetics: APE1493
 C:Superfamily: Aeropyrum pernix hypothetical protein APE1493

Query Match 0.7%; Score 8; DB 2; Length 240;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 GYSVSSML 439
 |||||
 DB 107 GYSVSSML 114

RESULT 29

A81345
 histidine-binding protein precursor Cj0734c (imported) - Campylobacter jejuni (strain NC8)
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: A81345
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barral Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypoxanthine auxotrophy
 A:Reference number: A81250; MUID:20150912; PMID:10688204
 A:Accession: A81345
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-251 <PAR>
 A:Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CA873008.1; PID:G696819
 A:Experimental source: serotype O2, strain NCTC 11168
 A:Genetics: Cj0734c
 C:Superfamily: lysine-arginine-ornithine-binding protein

Query Match 0.7%; Score 8; DB 2; Length 251;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 922 GTLGNTA 929
 |||||
 DB 146 GTLGNTA 153

RESULT 30

B55904
 C 3.4.25.1 proteasome endopeptidase complex () chain PR6 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: proteasome alpha chain; proteasome PR6; protein O2065; protein YOL03
 C:Species: Saccharomyces cerevisiae
 C:Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 03-Jun-2002
 C:Accession: B55904; S66721; S66723; S47910
 R:Heinemeyer, W.; Treondle, N.; Albrecht, G.; Wolf, D.H. Biochemistry 33, 12229-12237, 1994
 A:Title: PR6 and PR6, the last missing genes encoding 20S proteasome subunits from yeast
 A:Reference number: B55904; MUID:95001940; PMID:7918444
 A:Accession: B55904
 A:Status: preliminary
 A:Molecule type: DNA

```

A:Residues: 1-254 <HEI>
A:Cross-references: GB:L34348; NID:G511904; PIDN:AAA4903.1; PID:G511905
R:Habib, B.; Hattenberg, U.; Hollenberg, C.P.; Ramezani Rad, M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66703
A:Accession: S66721
A:Molecule type: DNA
A:Residues: 1-254 <HAB>
A:Cross-references: EMBL:Z74780; NID:G1419832; PIDN:CAA9040.1; PID:G1419833; MIPS:YOL03
A:Experimental source: strain S288C
R:Angers, W.; Bens, V.; Rechmann, S.; Schwager, C.; Teodoru, C.; Voss, H.; Wiemann, S.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66723
A:Accession: S66723
A:Molecule type: DNA
A:Residues: 1-254 <ANS>
A:Cross-references: EMBL:Z74780; NID:G1419832; PIDN:CAA9040.1; PID:G1419833; MIPS:YOL03
A:Experimental source: strain S288C
R:Heinmeyer, W.; Troendle, N.; Albrecht, G.; Wolf, D.H.
submitted to the EMBL Data Library, July 1994
A:Description: PRE5 and PRE6, the last missing genes encoding 20S proteasome subunits fr
A:Reference number: S47908
A:Accession: S47910
A:Molecule type: DNA
A:Residues: 1-254 <HEM>
A:Cross-references: EMBL:U34348; NID:G511904; PIDN:AAA4903.1; PID:G511905
C:Genetics:
A:Gene: SGD:PRE6
A:Cross-references: SGD:S0005398; MIPS:YOL038w
A:Map position: 15L
C:Superfamily: multicatalytic endopeptidase complex chain C9
C:Keywords: hydrolase; proteinase

Query Match          0.7%; Score 8; DB 2; Length 254;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1109 SLEEVQQT 1116
DB      198 SLEEVQQT 205

RESULT 31
117985
proteasome component SPBC106.16 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T37985; T39273
R:Murphy, U.; Harris, D.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z21759
A:Accession: T37985
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-259 <MUR>
A:Cross-references: EMBL:Z69909; PIDN:CAA93796.1; GSPDB:GN00066; SPDB:SPAC19G10.14C
A:Experimental source: strain 972h-; cosmid c19G10
R:Lyme, M.; Rajandream, M.A.; Barrett, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21840
A:Accession: T39273
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-259 <LYN>
A:Cross-references: EMBL:AL110295; PIDN:CAE53732.1; GSPDB:GN00067; SPDB:SPBC106.16
A:Experimental source: strain 972h-; cosmid c106
C:Genetics: <MUR1>
A:Gene: SPAC19G10.14C
A:Map position: 1
A:Introns: 33/3; 158/2
C:Genetics: <LYN1>
A:Gene: SPBC106.16
A:Map position: 2

```

```

C:Superfamily: multicatalytic endopeptidase complex chain C9

Query Match          0.7%; Score 8; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1109 SLEEVQQT 1116
DB      194 SLEEVQQT 201

RESULT 32
S22090
catechol O-methyltransferase (EC 2.1.1.6) - rat
N:Alternate names: catechol-O-methyltransferase
C:Species: Rattus norvegicus (Norway rat)
C>Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: S22090; JQ0787; I52356
R:Tenhunen, J.; Salminen, M.; Jalanko, A.; Ukkonen, S.; Uimannen, I.
submitted to the EMBL Data Library, June 1992
A:Description: Structure of the rat catechol-O-methyltransferase gene: Separate promoter
A:Reference number: S22090
A:Accession: S22090
A:Molecule type: DNA
A:Residues: 1-264 <TEN>
A:Cross-references: EMBL:Z12651; NID:G55889; PIDN:CAA78276.1; PID:G55890
R:Salminen, M.; Lundstroem, K.; Tilgmann, C.; Savolainen, R.; Kalkkinnen, N.; Uimannen, I.
Gene 93, 241-247, 1990
A>Title: Molecular cloning and characterization of rat liver catechol-O-methyltransferas
A:Reference number: JQ0787; MUID:91033034; PMID:2227437
A:Accession: JQ0787
A:Molecule type: mRNA
A:Residues: 44-264 <SAL>
A:Cross-references: GB:M60753; NID:G203336; PIDN:AAA40881.1; PID:G203337
A:Experimental source: liver
R:Tenhunen, J.; Uimannen, I.
Biochem. J. 296, 595-600, 1993
A>Title: Production of rat soluble and membrane-bound catechol O-methyltransferase forms
A:Reference number: I52356; MUID:94107221; PMID:8280056
A:Accession: I52356
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <RES>
A:Cross-references: EMBL:Z12651; NID:G55889; PIDN:CAA78276.1; PID:G55890
C:Comment: This enzyme catalyzes the transfer of a methyl group from S-adenosyl-L-methio
C:Genetics:
A:Gene: COMT
C:Superfamily: caffeineyl-CoA 3-O-methyltransferase
C:Keywords: methyltransferase; S-adenosylmethionine; transmembrane protein

Query Match          0.7%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 19;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1133 GLLLLALL 1140
DB      9 GLLLLALL 16

RESULT 33
S18541
hypothetical protein 3 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Oct-1999
C:Accession: S18541
R:Caillero, J.L.; Martinez, E.; Malpartida, F.; Hopwood, D.A.
Mol. Gen. Genet. 230, 401-412, 1991
A>Title: Organisation and functions of the actva region of the actinorhodin biosynthetic
A:Reference number: S18539; MUID:92114870; PMID:1766437
A:Accession: S18541
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <CAB>

```

A:Cross-references: EMBL:X85833; NID:g46812; PIDN:CAA41639.1; PID:g46815

Query Match 0.7%; Score 8; DB 2; Length 281;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 TAGARAAAF 715

DB 89 TAGARAAAF 96

RESULT 34

F71556

probable had hydrolase/phosphatase - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999

C:Accession: F71556

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis

A:Reference number: A71570; MUID:99000809; PMID:9784136

A:Accession: F71556

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <ARN>

A:Cross-references: GB:AE001284; GB:AE001273; NID:g3328494; PIDN:AAC67694.1; PID:g332850

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: CT103

Query Match 0.7%; Score 8; DB 2; Length 288;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PEARLTV 272

DB 141 PEARLTV 148

RESULT 35

C65217

allose kinase (EC 2.7.1.55) - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: C65217

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A:Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: C65217

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-309 <BLAT>

A:Cross-references: GB:AE000482; GB:U00096; NID:g2367349; PIDN:AACT7045.1; PID:g1790522;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: Y1CT

C:Superfamily: conserved hypothetical protein HI0182; glucose kinase homology

C:Keywords: phosphotransferase

Query Match 0.7%; Score 8; DB 2; Length 309;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 632 GAQGAAIL 639

DB 288 GAQGAAIL 295

RESULT 36

A85016

hypothetical protein AT901220 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: A85016

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617158

A:Accession: A85016

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <STO>

A:Cross-references: GB:NC_001268; NID:g7267619; PIDN:CAB80931.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT901220

A:Map position: 4

Query Match 0.7%; Score 8; DB 2; Length 360;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141

DB 44 LLLALLLV 51

RESULT 37

S19172

cytochrome P450 2B4 - rat (fragments)

N:Alternate names: cytochrome P450 LM2

N:Contains: oxidoreductase (EC 1.-.-.-)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 22-Nov-1993 #sequence_revision 21-Jul-1995 #text_change 04-Mar-2000

C:Accession: S19172

R:Yuan, P.M.; Ryan, D.E.; Levin, W.; Shively, J.E.

Proc. Natl. Acad. Sci. U.S.A. 80, 1169-1173, 1983

A:Title: Identification and localization of amino acid substitutions between two phenols

A:Reference number: S19172; MUID:83144040; PMID:6572377

A:Accession: S19172

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-158;159-200;201-310;311-367 <YU>

A:Experimental source: strain Long-Evans

C:Genetics:

A:Gene: CYP2B4

C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology

C:Keywords: chromoprotein; heme; iron; metalloprotein; microsome; monooxygenase; oxidore-

F312/Binding site: heme iron (lys) (axial ligand) #status predicted

QY 1134 LLLALLLV 1141

DB 6 LLLALLLV 13

RESULT 38

T31771

hypothetical protein F59D6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000

C:Accession: T31771

R:Bradshaw, H.; Graves, T.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F59D6.

A:Reference number: Z21083

A:Accession: T31771

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-383 <BRA>

A:Cross-references: EMBL:AF016435; PIDN:AA65877.1; GSPDB:GN00023; CESP:F59D6.2

A:Experimental source: strain Bristol NZ; clone F59D6

C:Genetics:
A:Gene: CESP:FS9D6.2
A:Map position: 5
A:Introns: 72/1; 100/1; 262/3
C:Superfamily: pepsin

Query Match 0.7%; Score 8; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1136 LIALLVFC 1143
Db 8 LIALLVFC 15

RESULT 39
G83568
probable permease of ABC transporter PA0605 [imported] - Pseudomonas aeruginosa (strain
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83568
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: G83568
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <STO>
A:Cross-references: GB:AE004497; GB:AE004091; NID:g9946478; PIDN:AA03994.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0605

Query Match 0.7%; Score 8; DB 2; Length 415;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1131 LGGILLLA 1138
Db 388 LGGILLLA 395

RESULT 40
AC2275
hypothetical protein alr3754 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075453.1; PID:g1713288; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3754

Query Match 0.7%; Score 8; DB 2; Length 433;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 982 VSGLIISA 989
Db 247 VSGLIISA 254

RESULT 41
T01721
hypothetical protein A_IG002N01.8 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01721
R:Schneer, P.; Magri, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana IG002N01.
A:Reference number: Z14407
A:Accession: T01721
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-444 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191131
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 68/1; 235/3; 302/3; 374/3
A:Note: A_IG002N01.8

Query Match 0.7%; Score 8; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1134 LLLALLLV 1141
Db 115 LLLALLLV 122

RESULT 42
O4RTP
cytochrome P450 2B1 - rat
N:Alternate names: cytochrome P450 b; cytochrome P450, phenobarbital-inducible
N:Contains: unspecified monooxygenase (EC 1.14.14.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Aug-1982 #sequence_revision 17-May-1996 #text_change 03-Mar-2000
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.
Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797, 1982
A:Title: Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobar
A:Reference number: A93912; MUID:82222224; PMID:6953431
A:Accession: A00176
A:Molecule type: mRNA
A:Residues: 6'491 <FU>
A:Cross-references: EMBL:J00719; NID:G203752; PIDN:AAA1024.1; PID:G203753
A:Note: The authors translated the codon GAT for residue 166 as Glu, CTG for residue 292
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.
Proc. Natl. Acad. Sci. U.S.A. 79, 5443, 1982
A:Title: Primary structure of a cytochrome P450: coding nucleotide sequence of phenobar
A:Reference number: A93925
A:Accession: A93925
A:Contents: annotation
A:Note: The mistranslations shown in reference A93912 are acknowledged
R:Roberts, E.S.; Hopkins, N.E.; Zaluzec, E.U.; Gage, D.A.; Alworth, W.L.; Hollenberg, P.
Biochemistry 33, 3766-3771, 1994
A:Title: Identification of active-site peptides from (3)H-labeled 2-ethylnaphthalene-i
A:Reference number: A54251; MUID:94190899; PMID:8142377
A:Accession: A54251
A:Molecule type: Protein
A:Residues: 290-301, 'X' <ROB>
R:Suwa, Y.; Mizukami, Y.; Sogawa, K.; Fujii-Kuriyama, Y.
J. Biol. Chem. 260, 7980-7984, 1985
A:Title: Gene structure of a major form of phenobarbital-inducible cytochrome P-450 in r
A:Reference number: A22363; MUID:85234490; PMID:2989270
A:Accession: A22363
A:Molecule type: DNA
A:Residues: 1-91, 'P', 93-204, 'R', 206-327, 'V', 329-356, 'H', 358-391, 'R', 393-415, 'V', 417-433,
A:Cross-references: GB:L00320; NID:G203816; PIDN:AAA1046.1; PID:G203818
A:Note: The authors translated the codon CAG for residue 57 as Gly, CCT for residue 92 a
as Glu, AAA for residue 236 as Leu, ACG for residue 259 as Asn, GTT for residue 328 as I
as Arg

R:Rangarajan, P.N.; Ravishankar, H.; Padmanaban, G.
Biochem. Biophys. Res. Commun. 144, 258-263, 1987
A:Title: Isolation of a cytochrome P-450 gene variant and characterization of its 5' fl
A:Reference number: A29298; PMID:87213174; PMID:3579906
A:Accession: A29298
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-57 <RAN>
R:Oesch, F.; Waxman, D.J.; Morrissey, J.J.; Honisch, W.; Kissel, W.; Friedberg, T.
Arch. Biochem. Biophys. 270, 23-32, 1989
A:Title: Antibodies targeted against hypervariable and constant regions of cytochromes P
A:Reference number: S03854; PMID:89192373; PMID:2539047
A:Accession: S03854
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-18,146-160, 'E', 162-165,166,330-361,362-380,402-423 <OES>
R:Botelho, L.H.; Ryan, D.E.; Levin, W.
J. Biol. Chem. 254, 5635-5640, 1979
A:Title: Amino acid compositions and partial amino acid sequences of three highly purified
or 3-methylcholanthrene.
A:Reference number: A92255; PMID:79194111; PMID:109438
A:Accession: A92255
A:Molecule type: protein
A:Residues: 1-3, 'T', 5-22 <BOT>
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Taniguchi, T.; Muramatsu, M.
Int. Symp. Princess Takamatsu Cancer Res. Fund 12, 31-40, 1982
A:Title: Molecular cloning and coding nucleotide sequence of complementary DNA of cytoch
A:Reference number: I54796; PMID:83160754; PMID:6300027
A:Accession: I54796
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 6-491 <RES>
A:Cross-references: GB:M37134; NID:G203784; PIND:AAQ42028.1; PID:G203785
C:Genetics:
A:Gene: CYP2B1
A:Introns: 57/3; 112/1; 162/1; 215/3; 274/3; 322/1; 384/3; 432/1
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F:295-458/Domain: cytochrome P450 homology <P45>
F:302/Active site: Thr #status predicted
F:436/Binding site: heme iron (Cys) (axial ligand) #status predicted
Query Match 0.7%; Score 8; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1134 LLLALLV 1141
DB 6 LLLALLV 13
RESULT 43
OARTP2
cytochrome P450 2B2 - rat
N:Alternate names: cytochrome P450 PB-4; cytochrome P450, phenobarbital-inducible; cyto
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 04-Dec-1986 #sequence revision 17-May-1996 #text change 01-Dec-2000
C:Accession: A21162; A00177; B00176; B92255; S15589; A21872; A32736; S03855; I59060
R:Mizukami, Y.; Sogawa, K.; Suwa, Y.; Muramatsu, M.; Fujii-Kuriyama, Y.
Proc. Natl. Acad. Sci. U.S.A. 80, 3958-3962, 1983
A:Title: Gene structure of a phenobarbital-inducible cytochrome P-450 in rat liver.
A:Reference number: A21162; PMID:83247397; PMID:630654
A:Accession: A21162
A:Molecule type: DNA
A:Residues: 1-472, 'W', 474-491 <MTZ>
A:Cross-references: EMBL:J00728; NID:G203845; PIND:AAA41056.1; PID:G203847
A:Note: the authors translated the codon AGR for residue 4 as Thr, and ATG for residue 3
R:Freyl, A.B.; Waxman, D.J.; Kreibich, G.
J. Biol. Chem. 260, 15253-15255, 1985
A:Title: The structure of phenobarbital-inducible rat liver cytochrome P-450 isoenzyme P
A:Reference number: A00177; PMID:86059379; PMID:3877725
A:Accession: A00177

A:Molecule type: protein
A:Residues: 1-291, 'P', 293-320, 'AE', 323-475, 'D', 477-491 <FRE>
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.
Proc. Natl. Acad. Sci. U.S.A. 79, 2793-2797, 1982
A:Title: Primary structure of a cytochrome P-450: coding nucleotide sequence of phenobar
A:Reference number: A93912; PMID:82222224; PMID:695431
A:Accession: B00176
A:Molecule type: mRNA
A:Residues: 6-359, 'S', 361-362, 'V', 364-366, 'V', 368-406, 'S', 408-416, 'N', 418, 'A', 420-477, 'G
A:Note: nucleotide sequence for residues 1-5 is not given
A:Note: the authors translated the codon GAR for residue 166 as Glu, CTG for residue 292
R:Fujii-Kuriyama, Y.; Mizukami, Y.; Kawajiri, K.; Sogawa, K.; Muramatsu, M.
Proc. Natl. Acad. Sci. U.S.A. 79, 5443, 1982
A:Title: Primary structure of a cytochrome P450: coding nucleotide sequence of phenobar
A:Reference number: A93925
A:Contents: annotation; revisions
A:Note: the mistranslations in reference A93912 are acknowledged
R:Botelho, L.H.; Ryan, D.E.; Levin, W.
J. Biol. Chem. 254, 5635-5640, 1979
A:Title: Amino acid compositions and partial amino acid sequences of three highly purified
or 3-methylcholanthrene.
A:Reference number: A92255; PMID:79194111; PMID:109438
A:Accession: B92255
A:Molecule type: protein
A:Residues: 1-3, 'T', 5-22 <BOT>
R:Lacroix, D.; Desrochers, M.; Lambert, M.; Anderson, A.
Gene 86, 201-207, 1990
A:Title: Alternative splicing of mRNA encoding rat liver cytochrome P450e (P450IIB2).
A:Reference number: S15589; PMID:90215299; PMID:2325573
A:Accession: S15589
A:Molecule type: mRNA
A:Residues: 105-113, 'P', 115-274, 'VSPAMRE', 275-321, 'E', 323-491 <LAC>
A:Cross-references: EMBL:M34452; NID:G203679; PIND:AAA41004.1; PID:G203680
A:Note: translation of the nucleotide sequence is not complete
R:Phillips, I.R.; Shephard, E.A.; Ashworth, A.; Rabin, B.R.
Gene 24, 41-52, 1983
A:Reference number: A21872
A:Accession: A21872
A:Molecule type: mRNA
A:Residues: 168-321, 'E', 323-443, 'K', 445-491 <PHI>
R:Afloiter, M.; Anderson, A.
Biochem. Biophys. Res. Commun. 118, 655-662, 1984
A:Title: Segmental homologies in the coding and 3' non-coding sequences of rat liver cyt
A:Reference number: A32736; PMID:84153837; PMID:632758
A:Accession: A32736
A:Molecule type: mRNA
A:Residues: 385-491 <AFP>
A:Cross-references: GB:K01626; NID:G203782; PIND:AAA41037.1; PID:G203783
R:Oesch, F.; Waxman, D.J.; Morrissey, J.J.; Honisch, W.; Kissel, W.; Friedberg, T.
Arch. Biochem. Biophys. 270, 23-32, 1989
A:Title: Antibodies targeted against hypervariable and constant regions of cytochromes P
A:Reference number: S03854; PMID:89192373; PMID:2539047
A:Accession: S03854
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 323-431 <RES>
A:Cross-references: GB:M3234; NID:G203848; PIND:AAA41057.1; PID:G554434
C:Genetics:
A:Gene: CYP2B2
A:Introns: 384/3
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: alternative splicing; chromoprotein; electron transfer; endoplasmic reticul
F:295-458/Domain: cytochrome P450 homology <P45>
F:436/Binding site: heme iron (Cys) (axial ligand) #status predicted

```

Query Match      0.7%; Score 8; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1134 LLLALLLV 1141
        |||||
        6 LLLALLLV 13

RESULT 44
B31047
testosterone 16alpha-hydroxylase (EC 1.14.14.-) cytochrome P450 2B10 - mouse
N/Alternate names: cytochrome P450CB; cytochrome P450p3/46
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
C/Accession: B31047; A60559
R/Noshiro, M.; Lako, M.; Kawajiri, K.; Negishi, M.
Biochemistry 27, 6434-6443, 1988
A/Title: R1P locus: regulation of female-specific isozyme (I-P-450-16-alpha) of testosterone
A/Reference number: A31047; MUID:89118235; PMID:3219345
A/Accession: B31047
A/Molecule type: mRNA
A/Residues: 1-500 <NO2>
A/Cross-references: EMBL:M21856; NID:g201968; PIDN:AAA40425.1; PID:g201969
A/Note: the authors translated the codon TTA for residue 55 as Phe, AGA for residue 133
as Val, AAA for residue 236 as Asn, AGG for residue 251 as Lys, AAG for residue 253 as
3 as Asp
A/Note: the authors translated the codon GCA for residue 281 as Thr, TTC for residue 283
8 as Arg, ACA for residue 321 as Ala, CTG for residue 331 as Leu, TCA for residue 349 as
8 Ala, and GAC for residue 479 as Gly
A/Note: the sequence nucleotide translation from Fig. 8 is inconsistent with the nucleot
R/Bornheim, L.M.; Correia, M.A.
Mol. Pharmacol. 36, 377-383, 1989
A/Title: Purification and characterization of a mouse liver cytochrome P-450 induced by
A/Reference number: A60559; MUID:89384475; PMID:2779523
A/Accession: A60559
A/Molecule type: protein
A/Residues: 1-15 <BOR>
A/Note: this enzyme was induced by cannabidiol
C/Genetics:
A/Gene: CYP2b-10
C/Superfamily: human cytochrome P450 CYP2B; cytochrome P450 homology
C/Keywords: chromoprotein; electron transfer; endoplasmic reticulum; heme; iron; metallo
F/295-467/Domain: cytochrome P450 homology <P45>
F/445/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match      0.7%; Score 8; DB 2; Length 500;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1134 LLLALLLV 1141
        |||||
        6 LLLALLLV 13

RESULT 45
139311
Kruppel-type zinc finger protein ZNF74 - human
C/Species: Homo sapiens (man)
C/Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 01-Dec-2000
C/Accession: 139311; F42825; S37482
R/Aubry, M.; Demczuk, S.; Desmazet, C.; Alkem, M.; Auriat, A.; Julien, J.P.; Rouleau, G.A.
Hum. Mol. Genet. 2, 1583-1587, 1993
A/Title: Isolation of a zinc finger gene consistently deleted in DiGeorge syndrome.
A/Reference number: 139311; MUID:94093543; PMID:8268910
A/Accession: 139311
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-572 <RES>
A/Cross-references: EMBL:X71623; NID:G1050529; PIDN:CAA50632.1; PID:G1050530
A/Note: submitted to the EMBL Data Library, April 1993, revised 01-NOV-1995
R/Aubry, M.; Maineau, C.; Zhang, F.R.; Zahed, L.; Figlewicz, D.; Delattre, O.; Thomas,
Genomics 13, 641-648, 1992

```

```

A/Title: Cloning of six new genes with zinc finger motifs mapping to short and long arms
A/Reference number: A42825; MUID:92347859; PMID:1639391
A/Accession: F42825
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: nucleic acid
A/Residues: 427-444, 'V', 446-511 <AUB>
A/Note: sequence extracted from NCBI Backbone (NCBI:109775)
C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match      0.7%; Score 8; DB 2; Length 572;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      898 FSCSLLS 905
        |||||
        269 FSCSLLS 276

RESULT 46
136917
glycoprotein IIB - yellow baboon (fragment)
C/Species: Papio cynocephalus, Papio hamadryas cynocephalus (yellow baboon)
C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 29-Sep-1999
C/Accession: 136917
R/Hayzer, D.C.; Shoji, M.; Kim, T.M.; Runge, M.S.; Hanson, S.R.
Gene 151, 267-271, 1994
A/Title: Alternative splicing of the mRNA encoding baboon glycoprotein receptor GPIIB.
A/Reference number: 136916; MUID:95129873; PMID:7828888
A/Accession: 136917
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-604 <RES>
A/Cross-references: GB:L12233; NID:G176584; PIDN:AA65936.1; PID:G176585
C/Superfamily: integrin alpha-2b chain

Query Match      0.7%; Score 8; DB 2; Length 604;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1130 VLGGLLLL 1137
        |||||
        569 VLGGLLLL 576

RESULT 47
S60658
legumin - Gnetum gnemon
N/Alternate names: IIS globulin; seed storage protein
C/Species: Gnetum gnemon
C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Aug-1999
C/Accession: S60658
R/Braun, H.; Horstmann, C.; Baunlein, H.
submitted to the EMBL Data Library, August 1995
A/Description: Legumins of the Gnetatae: characterization and evolutionary relationships
A/Reference number: S60658
A/Accession: S60658
A/Molecule type: mRNA
A/Residues: 1-607 <BRA>
A/Cross-references: EMBL:Z50779; NID:94949870; PIDN:CAA00642.1; PID:94949871
C/Superfamily: glycinin
C/Keywords: seed; storage protein

Query Match      0.7%; Score 8; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1134 LLLALLLV 1141
        |||||
        10 LLLALLLV 17

RESULT 48
A69845

```

Na+/H+ antiporter homolog yjbQ - *Bacillus subtilis*
 C/Species: *Bacillus subtilis*
 C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C/Accession: A69845
 R/Kunitz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertozzi, C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chik, A.; Ertlich, S.D.; Emerson, P.T.; Entian, K.D.; Erington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Geller, K.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, S.; Sadate, Y.; Sato, T.; Scanlon, A.; Schleich, S.; Schreier, R.; Scoffone, F.; Seliguchi, J.; Sekowska, A.; Serot, A.; Whitters, P.; Wapat, A.; Yamamoto, H.; Yamane, K.; Yamamoto, K.; Yata, K.; Yoshida, K.
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A/Reference number: A69580; MUID:98044033; PMID:9384377
 A/Accession: A69845
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-614 <KUN>
 A/Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CAB13021.1; PID:el183184;
 A/Experimental source: strain 168
 C/Genetics:
 A/Gene: yjbQ

Query Match 0.7%; Score 8; DB 2; Length 614;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 |||||
 DB 313 LLLALLV 320

RESULT 49
 AC0527
 Enterichrome protein protein FhuB precursor STY0221 [imported] - *Salmonella enterica* su
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A/Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C/Accession: AC0527
 R.Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
 A/Reference number: AB0502; PMID:11677608
 A/Accession: AC0527
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-685 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD01355.1; PID:gl6501482; GSPDB:GN00176
 C/Genetics:
 A/Gene: STY0221
 C/Superfamily: vitamin B12 transport protein btuC

Query Match 0.7%; Score 8; DB 2; Length 685;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 |||||
 DB 36 LLLALLV 43

RESULT 50
 G96661
 hypothetical protein F24D7.13 [imported] - *Arabidopsis thaliana*

C/Species: *Arabidopsis thaliana* (mouse-ear cress)
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C/Accession: G96661
 R/Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A/Reference number: A66141; MUID:21016719; PMID:11130712
 A/Accession: G96661
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-767 <STO>
 A/Cross-references: GB:AE005173; NID:95456164; PIDN:AA09152.1; GSPDB:GN00141
 C/Genetics:
 A/Gene: F24D7.13
 A/Map position: 1

Query Match 0.7%; Score 8; DB 2; Length 767;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 AVGAPLED 575
 |||||
 DB 559 AVGAPLED 566

RESULT 51
 EB3349
 Probable ClpA/B-type proteinase PA2371 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C/Species: *Pseudomonas aeruginosa*
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: EB3349
 R.Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mlczoguchi, S.D.; Warenaer, P.; Hickey, M.J.; Br adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lathig, K.; Lim, .; Loty, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
 A/Reference number: AB2950; MUID:20437337; PMID:10984043
 A/Accession: EB3349
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-849 <STO>
 A/Cross-references: GB:AE004663; GB:AE004091; NID:99948405; PIDN:AA05759.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: PA2371
 C/Superfamily: endopeptidase Clp ATP-binding chain

Query Match 0.7%; Score 8; DB 2; Length 849;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALLL 1140
 |||||
 DB 111 GLLALLL 118

RESULT 52
 J01674
 protein kinase TMK1 (EC 2.7.1.-), receptor type precursor - *Arabidopsis thaliana*
 C/Species: *Arabidopsis thaliana* (mouse-ear cress)
 C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jun-1999
 C/Accession: J01674
 R.Chang, C.; Schaller, G.E.; Patterson, S.E.; Kwok, S.F.; Meyerowitz, E.M.; Blecker, A.
 Plant Cell 4, 1263-1271, 1992
 A/Title: The TMK1 gene from *Arabidopsis* codes for a protein with structural and biochemi
 A/Reference number: J01674; MUID:93076110; PMID:1332795

A:Accession: J01674
 A:Molecule type: DNA
 A:Residues: 1-942 <CHA>
 A:Cross-references: GB:L00670; NID:G16687; PIDN:AAA2876.1; PID:G16688
 C:Superfamily: protein kinase Xa1; leucine-rich alpha-2-glycoprotein repeat homology; F
 C:Keywords: ATP, autophosphorylation; glycine-rich; phosphotransferase; receptor; serine
 F:1-22/Domain: signal sequence #status predicted <Sig>
 F:23-942/Product: protein kinase TMX1 #status predicted <MAT>
 F:65-88/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
 F:89-111/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
 F:112-136/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
 F:136-160/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
 F:161-186/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
 F:187-203/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
 F:210-233/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>
 F:233-255/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>
 F:256-279/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>
 F:280-299/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>
 F:300-323/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>
 F:324-346/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>
 F:363-386/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>
 F:387-410/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>
 F:411-434/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>
 F:480-503/Domain: transmembrane #status predicted <TMH>
 F:586-872/Domain: protein kinase homology <KIN>
 F:594-602/Region: protein kinase ATP-binding motif
 F:86,99,158,164,171,363,533,587/Binding site: carbohydrate (Asn) (covalent) #status pred
 F:616,634,717,719/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 0.7%; Score 8; DB 1; Length 942;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1128 GSVGLGL 1135
 Db 486 GSVGLGL 493

RESULT 53
 B60017
 Outer capsid protein VP2 - bluetongue virus (serotype 3, strain South Africa-VACC)
 C:Species: bluetongue virus
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 16-Jul-1999
 C:Accession: B60017
 R:Gould, A.R.; Pritchard, L.I.
 Virus Res. 17, 31-52, 1990
 A:Title: Relationships amongst bluetongue viruses revealed by comparisons of capsid and
 A:Reference number: A60017; MUID:91021485; PMID:2171239
 A:Accession: B60017
 A:Molecule type: genomic RNA
 A:Residues: 1-959 <GOU>
 A:Cross-references: GB:X55801; NID:G297130; PIDN:CAA39323.1; PID:G297131
 C:Genetics:
 A:Map position: segment 2
 C:Superfamily: bluetongue virus VP2 protein
 C:Keywords: capsid protein; glycoprotein
 F:749,910/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 0.7%; Score 8; DB 1; Length 959;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1159 EKREKLE 1166
 Db 798 EKREKLE 805

RESULT 54
 T15446
 Hypothetical protein C07G1.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15446

R:Hawkins, J.
 submitted to the EMBL Data Library, May 1996
 A:Description: The sequence of C. elegans cosmid C07G1.
 A:Reference number: Z18352
 A:Accession: T15446
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-969 <HAM>
 A:Cross-references: EMBL:U58751; NID:G1326379; PID:G1326381; PIDN:AB00657.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone C07G1
 C:Genetics:
 A:Gene: CESP.C07G1.4
 A:Map position: 4
 A:introns: 110/3; 149/2; 515/1; 564/1; 600/2; 690/2; 755/3; 843/3; 884/3; 940/2

Query Match 0.7%; Score 8; DB 2; Length 969;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 672 EAVCLTAA 679
 Db 9 EAVCLTAA 16

RESULT 55
 A60163
 glycoprotein IIB - rat
 C:Species: Rattus sp. (rat)
 C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 13-Sep-1998
 C:Accession: A60163; B60163
 R:Poncz, M.; Newman, P.J.
 Blood 75, 1282-1289, 1990
 A:Title: Analysis of rodent platelet glycoprotein IIB: evidence for evolutionarily conse
 A:Reference number: A60163; MUID:90181615; PMID:2310828
 A:Accession: A60163
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1037 <PON>
 A:Accession: B60163
 A:Molecule type: protein
 A:Residues: 32-50 <PON2>
 C:Comment: This protein is proteolytically processed into a heavy chain and a light chain
 C:Superfamily: integrin alpha-2b chain
 C:Keywords: calcium; disulfide bond; glycoprotein

Query Match 0.7%; Score 8; DB 2; Length 1037;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 VLGGILLL 1137
 Db 1003 VLGGILLL 1010

RESULT 56
 A34269
 Integrin alpha-2b chain precursor - human
 M:Alternate names: antigen CD41b; platelet adhesion receptor glycoprotein IIB; platelet
 C:Species: Homo sapiens (man)
 C:Date: 25-May-1990 #sequence_revision 23-May-1997 #text_change 29-Sep-1999
 C:Accession: A34269; A29522; I57461; A28937; S00268; A28581; A36430; B32528; A32
 R:Heidenreich, R.; Eisman, R.; Surrey, S.; Delgrosso, K.; Bennett, J.S.; Schwartz, E.; P
 Biochemistry 29, 1232-1244, 1990
 A:Title: Organization of the gene for platelet glycoprotein IIB.
 A:Reference number: A34269; MUID:90212612; PMID:2322558
 A:Accession: A34269
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-286 'VP', 289-731 'AI', 734-1039 '<HEI>
 A:Cross-references: GB:M33319; NID:G183506; GB:M3320; NID:G183506; GB:M34334; NID:G1835
 A:Note: The authors translated the codon AAC for residue 819 as ARG, CGG for residue 931
 R:Poncz, M.; Eisman, R.; Heidenreich, R.; Silver, S.M.; Vialatte, G.; Surrey, S.; Schwart

J. Biol. Chem. 262, 8476-8482, 1987
 A.Title: Structure of the platelet membrane glycoprotein IIb. Homology to the alpha subu
 A.Reference number: A29522; PMID:87250457; PMID:2439501
 A.Accession: A29522
 A.Molecule type: mRNA
 A.Residues: 1-312, 'A', 314-327, 'H', 329-632, 'S', 634-1039 <PON>
 A.Cross-references: GB:J02764; NID:G190067; PIDN:AAA6014.1; PID:G190068
 A.Note: the codons given for residues 313-Gly (GCA) and 328-Asp (CAC) are inconsiste
 R:Frachet, P.; Uzan, G.; Thevenon, D.; Denarier, E.; Prandini, M.H.; Marguerite, G.
 Mol. Biol. Rep. 14, 27-33, 1990
 A.Title: GPIIb and GPIIb amino acid sequences deduced from human megakaryocyte cDNAs.
 A.Reference number: 157461; PMID:90265363; PMID:2345548
 A.Accession: 157461
 A.Status: preliminary; translated from GB/EMBL/DDJ
 A.Molecule type: mRNA
 A.Residues: 1-22, 'A', 24-312, 'A', 314-345, 'D', 347-564, 'D', 566-1028, 'HT', 1031-1039 <RES>
 A.Cross-references: GB:M34480; NID:G183510; PIDN:AAA5926.1; PID:G306794
 R:Prandini, M.H.; Denarier, E.; Frachet, P.; Uzan, G.; Marguerite, G.
 Biochem. Biophys. Res. Commun. 156, 595-603, 1988
 A.Title: Isolation of the human platelet glycoprotein IIb gene and characterization of t
 A.Reference number: A28937; PMID:89025907; PMID:2845986
 A.Accession: A28937
 A.Molecule type: DNA
 A.Residues: 1-62, 1021-1028, 'HT', 1031-1039 <PRA>
 A.Cross-references: GB:M22568; NID:G183448; PIDN:AAA52587.1; PID:G55310; GB:M22569; NID
 R:Uzan, G.; Frachet, P.; Lajmanovich, A.; Prandini, M.H.; Denarier, E.; Duperray, A.; Lc
 Eur. J. Biochem. 171, 87-93, 1988
 A.Title: CDNA clones for human platelet GPIIb corresponding to mRNA from megakaryocytes
 A.Reference number: S00268; PMID:88111709; PMID:3422188
 A.Accession: S00268
 A.Molecule type: mRNA
 A.Residues: 392-565, 'V', 567-728, 'E', 730-1039 <UZA>
 A.Cross-references: EMBL:X06831; NID:G35519; PIDN:CAA29987.1; PID:G35520
 A.Note: the authors translated the codon AGC for residue 24 as Tyr; the sequence from Fi
 R:Bray, P.F.; Rosa, J.P.; Johnston, G.I.; Shiu, D.T.; Cook, R.G.; Lau, C.; Kan, Y.W.; Mc
 J. Clin. Invest. 80, 1812-1817, 1987
 A.Title: Platelet glycoprotein IIb. Chromosomal localization and tissue expression.
 A.Reference number: A28581; PMID:88059639; PMID:3479442
 A.Accession: A28581
 A.Molecule type: mRNA
 A.Residues: 868-1039 <BRA1>
 A.Cross-references: GB:M18085; NID:G183503; PIDN:AAA52597.1; PID:G183504
 R:Bray, P.F.; Leung, C.S.I.; Shuman, M.A.
 J. Biol. Chem. 265, 9587-9590, 1990
 A.Title: Human platelets and megakaryocytes contain alternately spliced glycoprotein IIb
 A.Reference number: A36430; PMID:90277633; PMID:2351656
 A.Accession: A36430
 A.Molecule type: mRNA
 A.Residues: 888-1039 <BRA2>
 A.Cross-references: GB:J05494
 A.Accession: B36430
 A.Molecule type: mRNA
 A.Residues: 888-947, 982-1039 <BRA3>
 A.Cross-references: GB:J05494
 A.Note: this form was found in platelets, megakaryocytes, and an erythroleukemia line, B
 R:Hizawa, A.; Matsukage, A.; Shiku, H.; Takahashi, T.; Naito, K.; Yamada, K.
 Blood 69, 560-564, 1987
 A.Title: Purification and partial amino acid sequence of human platelet membrane glycopr
 A.Reference number: A90731; PMID:87101510; PMID:3801670
 A.Accession: B32528
 A.Molecule type: protein
 A.Residues: 487-501 <HR1>
 A.Accession: A32528
 A.Molecule type: protein
 A.Residues: 'X', 1027, 'X', 1029, 'X', 1031-1038 <HR2>
 R:Calvete, J.J.; Henschen, A.; Gonzalez-Rodriguez, J.
 Biochem. J. 261, 561-568, 1989
 A.Title: Complete localization of the intrachain disulphide bonds and the N-glycosylatio
 A.Reference number: S05249; PMID:89374157; PMID:2775232
 A.Accession: S05249
 A.Molecule type: protein
 A.Residues: 32-41, 64-66, 73-79, 91-95, 'X', 97, 109-116, 123-129, 156-158, 160, 'X', 162-165, 172-1
 621, 667-690, 692-701, 709-710, 'X', 712-717, 'X', 725-730, 735-744 <CAL1>

A.Accession: S05254
 A.Molecule type: protein
 A.Residues: 903-910, 'X', 912, 1019-1024 <CAL2>
 R:Calvete, J.J.; Alvarez, M.V.; Rivas, G.; Hew, C.L.; Henschen, A.; Gonzalez-Rodriguez,
 Biochem. J. 261, 551-560, 1989
 A.Title: Interchain and intrachain disulphide bonds in human platelet glycoprotein IIb.
 A.Reference number: S05248; PMID:89374156; PMID:2476117
 A.Accession: S05248
 A.Molecule type: protein
 A.Residues: 903-922, 934-939 <CAL3>
 R:Calvete, J.J.; Schaefer, W.; Henschen, A.; Gonzalez-Rodriguez, J.
 FEBS Lett. 272, 37-40, 1990
 A.Title: Characterization of the beta-chain N-terminus heterogeneity and the alpha-chain
 A.Reference number: S12871; PMID:91032137; PMID:2226834
 A.Contents: annotation; chemical and mass spectrometric analyses
 R:Calvete, J.J.; Muniz-Diaz, E.
 FEBS Lett. 328, 30-34, 1993
 A.Title: Localization of an O-glycosylation site in the alpha-subunit of the human plate
 A.Reference number: S35059; PMID:93345693; PMID:7688323
 A.Accession: S35059
 A.Status: preliminary
 A.Molecule type: protein
 A.Residues: 735-745, 751-765, 766-774, 775-782, 783-795, 801-810, 811-820, 868-877, 'X', 879-881
 R:Calvete, J.J.; Parmentier, S.; Leung, L.L.; McGreggor, J.L.
 Biochem. J. 279, 419-425, 1991
 A.Title: Separation of important new platelet glycoproteins (GP1a, GP1c, GP1c*, GP1a an
 A.Reference number: A56793; PMID:92061944; PMID:1953640
 A.Accession: A56793
 A.Molecule type: protein
 A.Residues: 32-42 <CAT>
 A.Experimental source: platelet
 R:Kofuts, J.C.; Plow, E.F.; Frelinger III, A.L.; D'Souza, S.E.; Dixon, D.; Lacy, J.; Sor
 Proc. Natl. Acad. Sci. U.S.A. 84, 7114-7118, 1987
 A.Title: Molecular cloning and chemical synthesis of a region of platelet glycoprotein I
 A.Reference number: A28411; PMID:88041058; PMID:2444974
 A.Accession: A28411
 A.Molecule type: mRNA
 A.Residues: 'REYDLQVQWLPVPIVLEAKAGRSPEVRSRSLALNP', 786-819, 'T', 821-853, 'A', 855-964, 'F
 A.Cross-references: GB:J02963; NID:G190069; PIDN:AAA60115.1; PID:G387027
 A.Experimental source: HEL erythroleukemia cell library
 A.Note: the 5' end of this mRNA contains Alu-like sequence shown by antibodies to synthe
 R:Kakogorenko, E.M.; Yakubenko, V.P.; Ingham, K.C.; Medved, L.V.
 Eur. J. Biochem. 237, 205-211, 1996
 A.Title: Thermal stability of individual domains in platelet glycoprotein IIbIIa.
 A.Reference number: S65437; PMID:96203926; PMID:8620874
 A.Accession: S65437
 A.Molecule type: protein
 A.Residues: 872 <MAK>
 A.Accession: S65438
 A.Molecule type: protein
 A.Residues: 32 <MKO>
 C.Comment: Platelet membrane glycoprotein IIb-IIa (CD41/CD61 heterodimer) is a receptor
 C.Genetics: GDB:ITGA2B; GR2B
 A.Gene: ITGA2B; GR2B
 A.Cross-references: GDB:120012; OMIM:273800
 A.Map position: 17q21.32-17q21.32
 C.Superfamily: integrin alpha-2b chain
 C.Keywords: alternative splicing; glycoprotein; heterodimer; platelet; pyroglyutamic acid
 F:1-31/Domain: signal sequence #status predicted <SIG>
 F:32-887/Product: platelet glycoprotein IIb alpha chain #status experimental <MAT>
 F:891-1039/Product: platelet glycoprotein IIb beta chain (form 1) #status experimental <
 F:903-1039/Product: platelet glycoprotein IIb beta chain (form 2) #status experimental <
 F:994-1018/Domain: transmembrane #status predicted <TM>
 F:1019-1039/Domain: intracellular #status predicted <CYT>
 F:46-280, 601, 963/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:87-96, 138-161, 177-198, 504-515, 521-576, 633-639, 705-718/Disulfide bonds: #status predict
 F:711/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:857-911, 916-921/Disulfide bonds: #status experimental
 F:878/Binding site: carbohydrate (Ser) (covalent) #status experimental
 F:891/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime
 Query Match 0.7%; Score 8; DB 2; Length 1039;
 Best Local Similarity 100.0%; Pred. No. 60;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VLGGLL 1137
|||||

Db 1004 VLGGLL 1011

RESULT 57
S40311
integrit - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #ext_change 21-Jul-2000
C.Accession: S40311
R.Wehli, M.; Diantonio, A.; Fearney, I.M.; Smith, R.J.; Wilcox, M.
Mech. Dev. 43, 21-36, 1993
A.Title: Cloning and characterization of alpha(P51), a novel Drosophila melanogaster int
A.Reference number: S40311; MUID:94059764; PMID:8240969
A.Accession: S40311
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-1146 <MEH>
A.Cross-references: EMBL:X73975; NID:g440143; PIDN:CAA52155.1; PID:g440144
C.Genetics:
A.Gene: FlyBase:mew
A.Cross-references: FlyBase:FBgn0004456
C.Superfamily: integrin alpha-2b chain

Query Match 0.7%; Score 8; DB 2; Length 1146;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||

Db 1094 GGLLLAL 1101

RESULT 58
RHMUB
cell surface glycoprotein CD11b precursor [validated] - human
M.Alternate names: complement receptor type 3 alpha chain; leukocyte adhesion protein Ma
eukocyte integrin alpha chain; neutrophil adherence receptor alpham chain
C.Species: Homo sapiens (man)
C.Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #ext_change 31-Dec-2000
C.Accession: A31108; A28915; A41600; A30892; A32218; A46526; A26091; 152567
R.Cori, A.L.; Kishimoto, T.K.; Miller, L.J.; Springer, T.A.
J. Biol. Chem. 263, 12403-12411, 1988
A.Title: The human leukocyte adhesion glycoprotein Mac-1 (complement receptor type 3, CD
B.
A.Reference number: A31108; MUID:88315033; PMID:2457584
A.Accession: A31108
A.Molecule type: mRNA
A.Residues: 1-1153 <COR>
A.Cross-references: GB:J03925; NID:g187284; PIDN:AAA59544.1; PID:g307148
A.Note: part of this sequence was confirmed by protein sequencing
R.Arnaout, M.A.; Gupta, S.K.; Pierce, M.W.; Tenen, D.G.
J. Cell Biol. 106, 2153-2158, 1988
A.Title: Amino acid sequence of the alpha subunit of human leukocyte adhesion receptor M
A.Reference number: A28915; MUID:88257215; PMID:2454331
A.Accession: A28915
A.Molecule type: mRNA
A.Residues: 1-499,501-965, 'P', 967-1153 <ARN>
A.Cross-references: GB:M18044; GB:J03270; GB:M19664; GB:X07421; NID:g186935; PIDN:AAA594
A.Note: the authors translated the codon TAG for residue 1129 as Thr
A.Note: part of this sequence, including the amino end of the mature protein, was confir
R.Shelley, C.S.; Arnaout, M.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 10525-10529, 1991
A.Title: The promoter of the CD11b gene directs myeloid-specific and developmentally reg
A.Reference number: A41600; MUID:92073318; PMID:1683702
A.Accession: A41600
A.Molecule type: DNA
A.Residues: 1-9 <SH>
A.Cross-references: GB:M76724; NID:g180018; PIDN:AAA58410.1; PID:g553215
R.Arnaout, M.A.; Remold-O'Donnell, E.; Pierce, M.W.; Harris, P.; Tenen, D.G.

Proc. Natl. Acad. Sci. U.S.A. 85, 2776-2780, 1988
A.Title: Molecular cloning of the alpha-subunit of human and guinea pig leukocyte adhesi
A.Reference number: A34193; MUID:88190151; PMID:2833753
A.Accession: A30892
A.Molecule type: mRNA
A.Residues: 917-1042 <AR2>
A.Cross-references: GB:M18044
R.Hickstein, D.D.; Hickey, M.J.; Ozols, J.; Baker, D.M.; Back, A.L.; Roth, G.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 257-261, 1989
A.Title: cDNA sequence for the alpha subunit of the human neutrophil adherence receptor
A.Reference number: A32218; MUID:89098893; PMID:2563162
A.Accession: A32218
A.Molecule type: mRNA
A.Residues: 9-1153 <HIC>
A.Cross-references: GB:J04145; NID:g189068; PIDN:AAA59903.1; PID:g386975
A.Note: part of this sequence was confirmed by protein sequencing
R.Pleming, J.C.; Pahl, H.L.; Gonzalez, D.A.; Smith, T.F.; Tenen, D.G.
J. Immunol. 150, 480-490, 1993
A.Title: Structural analysis of the CD11b gene and phylogenetic analysis of the alpha-1r
n during evolution.
A.Reference number: A46526; MUID:93123748; PMID:8419480
A.Accession: A46526
A.Status: not compared with conceptual translation
A.Molecule type: DNA
A.Residues: 9-1153 <ELE>
A.Cross-references: GB:S52227; NID:g263047; PIDN:AB24821.1; PID:g263049
A.Note: the last three bases of intron 13, CAG, are included in some but not all mature
R.Pierce, M.W.; Remold-O'Donnell, E.; Todd III, R.F.; Arnaout, M.A.
Biochem. Biophys. Acta 874, 368-371, 1986
A.Title: N-terminal sequence of human leukocyte glycoprotein Mol: conservation across si
A.Reference number: A50664; MUID:87076671; PMID:3539202
A.Accession: A26091
A.Molecule type: protein
A.Residues: 17-31 <PIE>
A.Experimental source: granulocytes
R.Pahl, H.L.; Rosmarin, A.G.; Tenen, D.G.
Blood 79, 865-870, 1992
A.Title: Characterization of the myeloid-specific CD11b promoter.
A.Reference number: 152567; MUID:92144986; PMID:1346576
A.Accession: 152567
A.Status: translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-9 <RES>
A.Cross-references: GB:M84477; NID:g180184; PIDN:AAA51960.1; PID:g553219
C.Comment: A common beta chain (CD18) forms a heterodimer with this chain to form Mac-1
C.Genetics:
A.Gene: GDB:ITGM; CR3A
A.Cross-references: GDB:120599; OMTM:120980
A.Map position: 16p11.2-16p11.2
A.Note: promoter contains a GATA motif and two Sp1 consensus binding sites
C.Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homo
C.Keywords: alternative splicing; calcium; cell adhesion; glycoprotein; heterodimer; mag
F:17-1153/Domain: signal sequence #status predicted <SIG>
F:17-1108/Domain: cell surface glycoprotein CD11b #status experimental <MAT>
F:148-318/Domain: extracellular #status predicted <EXT>
F:148-318/Domain: von Willebrand factor type A repeat homology <VWA2>
F:465-477/Region: calcium/magnesium binding #status predicted
F:530-538/Region: calcium/magnesium binding #status predicted
F:593-601/Region: calcium/magnesium binding #status predicted
F:1109-1134/Domain: transmembrane #status predicted <TM>
F:1135-1153/Domain: intracellular #status predicted <INT>
F:86,240,391,469,693,697,735,802,881,901,912,941,947,979,994,1022,1045,1051,1076/Binding

Query Match 0.7%; Score 8; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||

Db 1116 GGLLLAL 1123

RESULT 59

RMHULC

cell surface glycoprotein CD11c precursor - human
N:Alternate names: leukocyte adhesion receptor p150,95 alpha chain

C:Species: Homo sapiens (man)
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999

C:Accession: A36584; A35543; S00864
R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 12750-12751, 1990

A:Reference number: A36584

A:Contents: erratum

A:Accession: A36584

A:Molecule type: DNA

A:Residues: 1-1163 <COR>

A:Note: this revision to the sequence from reference A35543 includes the carboxyl end

R:Corbi, A.L.; Garcia-Aguilar, J.; Springer, T.A.

J. Biol. Chem. 265, 2782-2788, 1990

A:Title: Genomic structure of an integrin alpha subunit, the leukocyte p150,95 molecule.

A:Reference number: A35543; MUID:90153906; PMID:2303426

A:Accession: A35543

A:Molecule type: DNA

A:Residues: 1-834 <CO2>

A:Note: this sequence has been revised in reference A36584

R:Corbi, A.L.; Miller, L.V.; O'Connor, K.; Larson, R.S.; Springer, T.A.

EMBO J. 6, 4023-4028, 1987

A:Title: CDNA cloning and complete primary structure of the alpha subunit of a leukocyte

A:Accession: S00864; MUID:8816645; PMID:3327687

A:Molecule type: mRNA

A:Residues: 1-755, 'L', 757-1163 <CO3>

A:Cross-references: GB:M81695; EMBL:Y00093; NID:g487829; PIDN:AA59180.1; PID:g487830

A:Note: part of this sequence was confirmed by protein sequencing

C:Comment: A common beta chain (CD18) forms a heterodimer with CD11b to form Mac-1 on my

C:Genetics:

A:Gene: GDB:ITGAX; CD11C

A:Cross-references: GDB:119758; OMIM:151510

A:Map position: 16p11.2-16p11.2

C:Superfamily: cell surface glycoprotein CD11b; von Willebrand factor type A repeat homoc

C:Keywords: calcium; cell adhesion; glycoprotein; heterodimer; magnesium; tandem repeat;

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1163/Product: cell surface glycoprotein CD11c #status predicted <MAT>

F:20-1107/Domain: extracellular #status predicted <EXT>

F:149-319/Domain: von Willebrand factor type A repeat homology <VWA4>

F:1108-1133/Domain: transmembrane #status predicted <TM>

F:1134-1163/Domain: intracellular #status predicted <INT>

F:61,89,392,697,735,899,939,1050/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 0.7%; Score 8; DB 1; Length 1163;

Best Local Similarity 100.0%; Pred. No. 66;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139

DB 1115 GGLLLALL 1122

RESULT 60

G84922
hypothetical protein At2g48050 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: G84922

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.L.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84922

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1500 <STO>

A:Cross-references: GB:AE002093; NID:g4249412; PIDN:AAD13709.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g48050

A:Map position: 2

Query Match 0.7%; Score 8; DB 2; Length 1500;

Best Local Similarity 100.0%; Pred. No. 81;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1065 LRLVHNE 1072

DB 1202 LRLVHNE 1209

RESULT 61

138414

transcription factor IIC, box B-binding chain - human

C:Species: Homo sapiens (man)

C:Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 05-Nov-1999

C:Accession: I38414

R:LeCelle, N.D.; Fehnestock, M.L.; Shen, Y.; Aebersold, R.; Berk, A.J.

Proc. Natl. Acad. Sci. U.S.A. 91, 1652-1656, 1994

A:Title: Human TFIIIC Box B-Binding Subunit.

A:Reference number: A31132; MUID:94173888; PMID:8127861

A:Accession: I38414

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2109 <RES>

A:Cross-references: EMBL:U02619; NID:g414932; PIDN:AA17985.1; PID:g442362

C:Genetics:

A:Gene: GDB:GTF3C1; TFIIIC

A:Cross-references: GDB:223512

A:Map position: 16p12.1-16p11.2

Query Match 0.7%; Score 8; DB 2; Length 2109;

Best Local Similarity 100.0%; Pred. No. 11e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 REPARLL 271

DB 170 REPARLL 177

RESULT 62

T13924

sdh protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000

C:Accession: T13924

R:Nguyen, D.N.; Liu, Y.; Litisky, M.L.; Reinke, R.

Submitted to the EMBL Data Library, February 1997

A:Description: Sidelick, a member of the immunoglobulin superfamily, is required for pat

A:Reference number: Z17809

A:Accession: T13924

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2222 <NCU>

A:Cross-references: EMBL:U88578; NID:g4099554; PID:g4099555; PIDN:AAD09632.1

C:Genetics:

A:Gene: sdh

A:Cross-references: FlyBase:FBgn0021764

Query Match 0.7%; Score 8; DB 2; Length 2222;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 970 TTLRVONT 977

DB 1842 TTLRVONT 1849

RESULT 63

CGRU3A

collagen alpha 3 (VI) chain precursor [validated] - human

N.Contains: collagen alpha 3(VI) chain, splice form A9/N10(-)
 C/Species: Homo sapiens (man)
 C/Date: 21-Nov-1993 #sequence revision 12-Nov-1999 #text change 15-Sep-2000
 C/Accession: A59140; S13679; S24465; A57083; S28776; S00245; C31952; C29848; S26510; S48
 R.Chu, M.L.
 submitted to GenBank, May 1998
 A/Reference number: A59140
 A/Accession: A59140
 A/Status: not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-3176 <CHU>
 A/Cross-references: GB:X52022; NID:G3127925; PIDN:CAA36267.1; PID:G3127926
 R.Chu, M.L.; Zhang, R.; Pan, T.; Stokes, D.; Conway, D.; Kuo, H.-J.; Glanville, R.; May
 EMBL J. 9, 385-393, 1990
 A/Title: Mosaic structure of globular domains in the human type VI collagen alpha-3 chain
 A/Reference number: S13679; MUID:90151612; PMID:1689238
 A/Accession: S13679
 A/Molecule type: mRNA
 A/Residues: 1-30,237-313, 'CWV', 318-322, 'AR', 326-1815, 'PD', 1818-1819, 'ID', 1822-3176 <CHS>
 A/Cross-references: EMBL:X52022; NID:G3127925
 A/Accession: S24465
 A/Molecule type: protein
 A/Residues: 574-585,965-973, 'X', 975-976,1306-1325,1361-1377,1381-1401,1473-1506, 'X', 1508
 -1962, 'X', 1964-1965,2018-2037,2374-2410,2445-2459,2466-2469, 'X', 2471-2474,2504-2508, 'X',
 R.Zanussi, S.; Dollana, R.; Segat, D.; Bonaldi, P.; Colombatti, A.
 J. Biol. Chem. 267, 24082-24089, 1992
 A/Title: The human type VI collagen gene. mRNA and protein variants of the alpha3 chain
 A/Reference number: S28776; MUID:93054780; PMID:1339440
 A/Accession: A57083
 A/Molecule type: DNA
 A/Residues: 310-328 <ZAN>
 A/Accession: S28776
 A/Molecule type: mRNA
 A/Residues: 32-126, 'AK', 129-136, 'L', 138-236 <ZAZ>
 A/Cross-references: GB:S49432; NID:G260296; PIDN:AA824261.1; PID:G260297
 R.Chu, M.L.; Mann, K.; Deutzmann, R.; Pribula-Conway, D.; Hsu-Chen, C.C.; Bernard, M.P.;
 Eur. J. Biochem. 168, 309-317, 1987
 A/Title: Characterization of three constituent chains of collagen type VI by peptide seq
 A/Reference number: S00245; MUID:88029444; PMID:366527
 A/Accession: S00245
 A/Molecule type: mRNA, protein
 A/Residues: 2024-2046,2092-2156, 'R', 2203-2208, 'X', 2210-2211, 'X', 2213-2227,2228-2251,2214
 A/Cross-references: GB:X06196; NID:G30055; PIDN:CAA29557.1; PID:G1335034
 A/Note: the mRNA portion of the sequence corresponds to residues 2092-2157
 R.Chu, M.L.; Conway, D.; Pan, T.; Baldwin, C.; Mann, K.; Deutzmann, R.; Timpl, R.
 J. Biol. Chem. 263, 18601-18606, 1988
 A/Title: Amino acid sequence of the triple-helical domain of human collagen type VI.
 A/Reference number: A31952; MUID:89066644; PMID:3198591
 A/Accession: C31952
 A/Molecule type: mRNA
 A/Residues: 2038-2373 <CH4>
 A/Cross-references: GB:J04211; GB:M30778
 A/Note: parts of this sequence were determined by protein sequencing
 R.Well, D.; Mattei, M.G.; Passagge, E.; Van Cong, N.; Pribula-Conway, D.; Mann, K.; Deutz
 Am. J. Hum. Genet. 42, 435-445, 1988
 A/Title: Cloning and chromosomal localization of human genes encoding the three chains c
 A/Reference number: A29848; MUID:88161046; PMID:3348212
 A/Accession: C29848
 A/Molecule type: mRNA
 A/Residues: 2092-2151 <WEI>
 A/Cross-references: GB:M27449; NID:G291919; PIDN:AAA52057.1; PID:G291920
 A/Note: part of this sequence was determined by protein sequencing
 R.Jander, R.; Rautenberg, J.; Glanville, R.W.
 Eur. J. Biochem. 133, 39-46, 1983
 A/Title: Further characterization of the three polypeptide chains of bovine and human sh
 A/Reference number: S26510; MUID:83209648; PMID:6852033
 A/Accession: S26510
 A/Molecule type: protein
 A/Residues: 'SAIVAGVAGV' <JAN>
 A/Note: this sequence cannot be reliably placed and probably represents the results from
 R.Mayer, U.; Poeschl, E.; Nischt, R.; Specks, U.; Pan, T.C.; Chu, M.L.; Timpl, R.
 Eur. J. Biochem. 225, 573-580, 1994

A/Title: Recombinant expression and properties of the Kunitz-type protease-inhibitor mod
 A/Reference number: S48709; MUID:95045506; PMID:7525281
 A/Accession: S48709
 A/Molecule type: mRNA
 A/Residues: 'NRAMIFFLCAGRAAA', 3102-3176 <MAY>
 A/Note: engineered sequence to allow isolation of the Kunitz-type proteinase inhibitor h
 R.Arnaud, B.; Merigau, K.; Salvidian, P.; Norris, F.; Norris, K.; Bjorn, S.; Olsen, O.;
 submitted to the Brookhaven Protein Data Bank, August 1994
 A/Reference number: A52812; PDB:1KXMT
 A/Contents: annotation; X-ray crystallography, 1.6 angstroms, residues 3106-3160
 A/Note: engineered sequence expressed in *Saccharomyces cerevisiae* strain mc-663
 C/Comment: Proline and lysines at the third position of the tripeptide repeating unit (
 shine are 5-hydroxylated and subsequently O-glycosylated.
 C/Comment: The fibronectin type III repeat homology domain may be released during proces
 C/Genetics:
 A/Genes: GDB:COL6A3
 A/Cross-references: GDB:119066; OMIM:120250
 A/Map position: 2q37.3-2q37.3
 C/Complex: type VI collagen is a heterotrimer of one alpha 1(VI) chain (see PIR:CGHUIA),
 associations among trimer amino- and carboxyl-terminal domains (with disulfide bonds).
 C/Function:
 A/Description: structural component of extracellular tissue microfibrils associated with
 C/Superfamily: collagen alpha 3(VI) chain; animal kunitz-type proteinase inhibitor homol
 C/Keywords: alternative splicing; blocked amino end; cell binding; coiled coil; extracel
 F/1-25/DNA: signal sequence #status predicted <SIG>
 F/1-25/DNA: Product: collagen alpha 3(VI) chain #status predicted <MAT1>
 F/26-3176/Product: amino-terminal nonhelical #status predicted <ANH>
 F/26-30,237-3176/Product: collagen alpha 3(VI) chain, splice form A9/N10(-) #status pred
 F/327-203/DNA: von Willebrand factor type A repeat homology <W01>
 F/240-405/DNA: von Willebrand factor type A repeat homology <W02>
 F/443-608/DNA: von Willebrand factor type A repeat homology <W03>
 F/637-802/DNA: von Willebrand factor type A repeat homology <W04>
 F/835-999/DNA: von Willebrand factor type A repeat homology <W05>
 F/1027-1191/DNA: von Willebrand factor type A repeat homology <W06>
 F/1123-1394/DNA: von Willebrand factor type A repeat homology <W07>
 F/1434-1599/DNA: von Willebrand factor type A repeat homology <W08>
 F/1637-1802/DNA: von Willebrand factor type A repeat homology <W09>
 F/1836-2005/DNA: von Willebrand factor type A repeat homology <W10>
 F/2038-2373/DNA: interrupted helical
 F/2040-2042/DNA: cell attachment (R-G-D) motif
 F/2136-2138/DNA: cell attachment (R-G-D) motif
 F/2148-2150/DNA: cell attachment (R-G-D) motif
 F/2154-2156/DNA: cell attachment (R-G-D) motif
 F/2370-2372/DNA: cell attachment (R-G-D) motif
 F/2374-3176/DNA: carboxyl-terminal nonhelical #status predicted <CNH>
 F/2400-2571/DNA: von Willebrand factor type A repeat homology <W11>
 F/2617-2800/DNA: von Willebrand factor type A repeat homology <W12>
 F/2865-2886/DNA: alanine/lysine/proline/threonine/valine-rich repeats
 F/2987-3072/DNA: fibronectin type III repeat homology <FN3>
 F/3111-3161/DNA: animal kunitz-type proteinase inhibitor homology <BPI>
 F/326/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F/108,116,202,251,2079,2331,2558,2677,2861,3036/Binding site: carbonylate (Asn) (covale
 F/2087/Disulfide bonds: interchain #status predicted
 F/2100,2206,2239,2316,2319/Modified site: 4-hydroxyproline (Pro) #status experimental
 F/2103,2209,2212,2322,2337/Binding site: carboxylate (Lys) (covalent) #status experimen
 F/3111-3161,3120-3144,3136-3157/Disulfide bonds: #status predicted
 Query Match 0.7%; Score 8; DB 1; Length 3176;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 202 IQVGLVQY 209
 DB 1676 IQVGLVQY 1683
 RESULT 64
 D81666
 hypochelical protein TC0766 [imported] - Chlamydia muridarum (strain Nigg)
 C/Species: Chlamydia muridarum, Chlamydia trachomatis Moyn
 C/Date: 31-Mar-2000 #sequence revision 31-Mar-2000 #text change 11-May-2000
 C/Accession: D81666

R.Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: D81666
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-47 <TEXT>
 A:Cross-references: GB:AE002345; GB:AE002160; NID:g7190791; PIDN:AAF39569.1; PID:g719079
 A:Experimental source: strain Ni99 (MoPn)
 A:Genetics: TC0766

Query Match
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALL 1140
 DB 4 LLLALL 10

RESULT 65
 C65069
 hypothetical protein b2858 - *Escherichia coli* (strain K-12)
 C:Species: *Escherichia coli*
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: C65069
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shaoh, Y.
 Science 277, 1453-1462, 1997
 A>Title: The complete genome sequence of *Escherichia coli* K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: C65069
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-73 <BLAT>
 A:Cross-references: GB:AE000369; GB:U00096; NID:g2367168; PIDN:AACT5897.1; PID:g1789222;
 A:Experimental source: strain K-12, substrain MG1655

Query Match
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EKKREK 1164
 DB 32 EKKREK 38

RESULT 66
 T10456
 dermaseptin B3 precursor - two-colored leaf frog
 C:Species: *Phyllomedusa bicolor* (two-colored leaf frog)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T10456
 R:Charpentier, S.; Amiche, M.; Mester, J.; Vouille, V.; Le Caer, J.P.; Niccolas, P.; Del J. Biol. Chem. 273, 14690-14697, 1998
 A>Title: Structure, synthesis, and molecular cloning of dermaseptins B, a family of skin A:Reference number: Z17027; MUID:9828974; PMID:9614066
 A:Accession: T10456
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-74 <CH>
 A:Cross-references: EMBL:Y16564; NID:g3256036; PIDN:CAA76288.1; PID:g3256037
 C:Superfamily: dermaseptin precursor; dermorphin precursor amino-terminal homology F.1.22/Domains: signal sequence #status predicted <SIG>
 F.23-74/Product: dermaseptin B3 #status predicted <MAT>

Query Match
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1157 EKKREK 1163
 DB 23 EKKREK 29

RESULT 67
 I51527
 integrin alpha 5 subunit - African clawed frog (fragment)
 C:Species: *Xenopus laevis* (African clawed frog)
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 29-Sep-1999
 C:Accession: I51527
 R:Whitaker, C.A.; Desimone, D.W.
 Development 117, 1239-1249, 1993
 A>Title: Integrin alpha subunit mRNAs are differentially expressed in early *Xenopus* embri
 A:Reference number: I51524; MUID:94008528; PMID:8404528
 A:Accession: I51527
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-76 <WHI>
 A:Cross-references: GB:U10191; NID:g214544; PIDN:AAA16249.1; PID:g214545
 C:Superfamily: integrin alpha-2b chain

Query Match
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 GRVYVYL 526
 DB 36 GRVYVYL 42

RESULT 68
 I51529
 integrin alpha V-like subunit - African clawed frog (fragment)
 C:Species: *Xenopus laevis* (African clawed frog)
 C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 29-Sep-1999
 C:Accession: I51529
 R:Whitaker, C.A.; Desimone, D.W.
 Development 117, 1239-1249, 1993
 A>Title: Integrin alpha subunit mRNAs are differentially expressed in early *Xenopus* embri
 A:Reference number: I51524; MUID:94008528; PMID:8404528
 A:Accession: I51529
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-77 <WHI>
 A:Cross-references: GB:U10190; NID:g214548; PIDN:AAA16251.1; PID:g214549
 C:Superfamily: integrin alpha-2b chain

Query Match
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 520 GRVYVYL 526
 DB 36 GRVYVYL 42

RESULT 69
 D70640
 hypothetical protein RV0689c - *Mycobacterium tuberculosis* (strain H37RV)
 C:Species: *Mycobacterium tuberculosis*
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: D70640
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: D70640
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-84 <COL>
A:Cross-references: GB:Z84395; GB:AJ123456; NID:g9261698; PIDN:CAB06452.1; PID:e293110;
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV0689c

Query Match 0.6%; Score 7; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1049 RCHLQOL 1055
DB 78 RCHLQOL 84

RESULT 70
G82776
ACP XF0672 (imported) - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 01-Dec-2000
C:Accession: G82776
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82776
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-85 <STM>
A:Cross-references: GB:AE003911; GB:AE003849; NID:g9105548; PIDN:AAF83482.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H
de-Nero, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.U.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kempner, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmetti, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saneili, R.V.; Sawabak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silva
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0672
C:Superfamily: acyl carrier protein; acyl carrier protein homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:8-79/Domain: acyl carrier protein homology <ACP>
F:43/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 0.6%; Score 7; DB 2; Length 85;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 414 MALDEF 420
DB 51 MALDEF 57

RESULT 71
B91093
hypothetical protein Ecs3714 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B91093
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B91093
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86 <NAV>
A:Cross-references: GB:BA000007; PIDN:BA837137.1; PID:g13363186; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: Ecs3714

Query Match 0.6%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EKEEEK 1164
DB 49 EKEEEK 55

RESULT 72
G85938
hypothetical protein Z4179 [imported] - Escherichia coli (strain O157:H7, substrain EDL5
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85938
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85938
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-86 <STO>
A:Cross-references: GB:AE005174; NID:g12517354; PIDN:AGS57971.1; GSPDB:GN00145; UMGF.Z41
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z4179

Query Match 0.6%; Score 7; DB 2; Length 86;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1158 EKEEEK 1164
DB 49 EKEEEK 55

RESULT 73
I49515
B144 protein B - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C:Accession: I49515
R:Tsuge, I.; Shen, F.
Immunogenetics 26, 378-380, 1987
A:Title: A gene in the H-2S: H-2D interval of the major histocompatibility complex which
A:Reference number: I49514; MUID:88031493; PMID:3117682
A:Accession: I49515
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-89 <RES>
A:Cross-references: GB:M18187; NID:g192097; PIDN:AAA37273.1; PID:g192099

Query Match 0.6%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGILL 1137
DB 36 LGGILL 42

RESULT 74

T03285
 anther-specific protein - rice
 C:Species: Oryza sativa (rice)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
 C:Accession: T03285
 R:Lee, J.Y.K.; Hodges, T.K.
 A:Submitted to the EMBL Data Library, July 1994
 A:Description: Genomic DNA sequence of a rice anther-specific gene.
 A:Reference number: Z14882
 A:Accession: T03285
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-94 <LEE>
 A:Cross-references: EMBL:U12171; NID:g607894; PID:g607895
 A:Experimental source: strain IR54
 C:Genetics:
 A:Gene: RRS2

Query Match 0.6%; Score 7; DB 2; Length 94;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 468 DCAVRVA 474
 |||||
 DB 31 DCAVRVA 37

RESULT 75

A64146
 YajC protein homolog H10241 - Haemophilus influenzae (strain Rd KW20)
 C:Species: Haemophilus influenzae
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
 C:Accession: A64146
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.; Glodek, A.; Kelley, J.M.; Weidman, J.
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: A64146
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-97 <TIGR>
 A:Cross-references: GB:U32710; GB:I42023; NID:g1573200; PIDN:AAC21909.1; PID:g1573206; T
 C:Superfamily: YajC protein
 C:Keywords: transmembrane protein
 F:8-24/Domain: transmembrane #status predicted <TMM>

Query Match 0.6%; Score 7; DB 1; Length 97;
 Best Local Similarity 100.0%; Pred. No. 89;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1055 LAKGTEV 1061
 |||||
 DB 42 LAKGTEV 48

Search completed: July 16, 2003, 08:34:15
 Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:28:30 ; Search time 52 Seconds
(without alignments)
4624.173 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167
Sequence: 1 MELPVTHTLPVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	1.1	1171	13	042094
2	11	0.9	191	6	029124
3	11	0.9	383	2	09R786
4	11	0.9	823	4	08WY18
5	10	0.9	780	13	006271
6	9	0.8	66	10	0943V3
7	9	0.8	84	16	0932R3
8	9	0.8	236	9	064016
9	9	0.8	236	16	032005
10	9	0.8	244	16	09RJL0
11	9	0.8	399	16	08XSH5
12	9	0.8	757	10	094LU9
13	9	0.8	1086	4	096H81
14	9	0.8	1160	11	09R200
15	9	0.8	1161	11	09WTV4
16	9	0.8	1187	13	098TF0

17	9	0.8	1196	13	098TF1	098TF1 cyprinus ca
18	8	0.7	59	4	09UR7	09ur7 homo sapien
19	8	0.7	77	9	09MC41	09mc41 bacterioph
20	8	0.7	97	4	000453	000453 homo sapien
21	8	0.7	103	16	08XEP2	08xep2 salmonella
22	8	0.7	105	2	09EY13	09ey13 klebsiella
23	8	0.7	112	11	08RIE9	08rie9 mus musculu
24	8	0.7	119	11	09OX14	09ox14 rattus norv
25	8	0.7	127	16	09RSB5	09rsb5 deinococcus
26	8	0.7	136	16	09KN21	09kn21 vibrio chol
27	8	0.7	141	11	061692	061692 mus musculu
28	8	0.7	146	13	08G930	08g930 bothrops in
29	8	0.7	169	16	08ZLM7	08zlm7 salmonella
30	8	0.7	169	16	08Z1W9	08z1w9 salmoneilla
31	8	0.7	181	13	090Y12	090y12 crocalus du
32	8	0.7	181	13	090Y11	090y11 crocalus du
33	8	0.7	185	16	097LC8	097lc8 clostridium
34	8	0.7	240	17	09YBV7	09ybv7 aeropyrum p
35	8	0.7	258	16	09RK95	09rk95 streptomyce
36	8	0.7	265	13	09PM56	09pm56 bothrops ja
37	8	0.7	265	13	08OC91	08oc91 bothrops in
38	8	0.7	269	5	09U168	09u168 leishmania
39	8	0.7	281	2	053905	053905 streptomyce
40	8	0.7	281	16	093122	093122 streptomyce
41	8	0.7	287	2	093QX2	093qx2 corynebacte
42	8	0.7	294	2	09XCW9	09xcw9 rhodobacter
43	8	0.7	298	16	084105	084105 chlamydia t
44	8	0.7	299	2	09R9K7	09r9k7 paracoccus
45	8	0.7	299	10	094K03	094k03 arabidopsis
46	8	0.7	309	2	005656	005656 bacillus ce
47	8	0.7	348	4	08TES5	08tes5 homo sapien
48	8	0.7	360	10	09M146	09m146 arabidopsis
49	8	0.7	383	5	016339	016339 caenorhabdi
50	8	0.7	385	5	09NFP1	09nfp1 leishmania
51	8	0.7	405	10	024416	024416 fragaria an
52	8	0.7	405	10	094FT6	094ft6 fragaria an
53	8	0.7	415	16	0915T3	0915t3 pseudomonas
54	8	0.7	433	16	08YOR5	08yor5 anabaena sp
55	8	0.7	444	10	004611	004611 arabidopsis
56	8	0.7	473	16	08XZT4	08xzt4 ralsionia s
57	8	0.7	491	2	09RE18	09re18 rhodobacter
58	8	0.7	491	2	09K559	09k559 rhodobacter
59	8	0.7	491	11	09WUD0	09wud0 mus musculu
60	8	0.7	491	11	064584	064584 rattus norv
61	8	0.7	497	13	073804	073804 fuqu rubrip
62	8	0.7	518	10	09A541	09a541 oryza sativ
63	8	0.7	573	4	08WY13	08wy13 homo sapien
64	8	0.7	577	16	08R6N2	08r6n2 thermocae
65	8	0.7	596	17	08TRM8	08trm8 methanosarc
66	8	0.7	607	10	039775	039775 gnetum gnet
67	8	0.7	614	16	031615	031615 bacillus su
68	8	0.7	622	13	091570	091570 xenopus lae
69	8	0.7	634	5	017795	017795 caenorhabdi
70	8	0.7	636	16	09CWV6	09cwv6 pasteurella
71	8	0.7	670	11	070491	070491 mus musculu
72	8	0.7	685	16	08Z9B5	08z9b5 salmonella
73	8	0.7	704	4	096PVI	096pvi homo sapien
74	8	0.7	739	12	09INI9	09ini9 kadipito vi
75	8	0.7	767	10	09CAD3	09cad3 arabidopsis
76	8	0.7	849	16	0911A8	0911a8 pseudomonas
77	8	0.7	952	10	08S920	08s920 oryza sativ
78	8	0.7	1036	6	09TU44	09tu44 canis famli
79	8	0.7	1036	6	09TUN8	09tun8 canis famli
80	8	0.7	1036	6	09TUN6	09tun6 sus scrofa
81	8	0.7	1049	5	08SYE1	08syel drosophila
82	8	0.7	1161	11	09OYF7	09oyef rattus norv
83	8	0.7	1500	10	09Z084	09z084 arabidopsis
84	8	0.7	2109	4	012789	012789 homo sapien
85	8	0.7	2221	5	09UIM1	09uim1 drosophila
86	8	0.7	2222	5	097394	097394 drosophila
87	8	0.7	4899	5	09VR91	09vr91 drosophila
88	7	0.6	22	6	09TRC4	09trc4 canis famli
89	7	0.6	25	4	09UL44	09ul44 homo sapien

90 7 0.6 47 16 Q9PJ22 chlamydia m
91 7 0.6 56 11 O09128 mus musculus
92 7 0.6 73 16 P76640 escherichia
93 7 0.6 77 13 O06276 xenopus lae
94 7 0.6 78 6 Q9N241
95 7 0.6 79 5 Q8S762
96 7 0.6 81 16 Q8Y0J1
97 7 0.6 89 11 O08844 s
98 7 0.6 84 15 O08844 mus musculus
99 7 0.6 84 15 O91182 human immun
100 7 0.6 85 16 P95035 mycobacteri
Q9PJ15 xyella fas

ALIGNMENTS

RESULT 1

O42094 PRELIMINARY; PRT; 1171 AA.
ID Q42094
AC Q42094
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DE 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
OS ALPHA1 integrin.
OC Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GIZZARD;
RX MEDLINE=97476270; PubMed=9334246;
RA Obara H., Hayashi K., Nishida W., Momiyama T., Uchida A., Ochi T.,
RA Sobue K.;
RT "Smooth muscle cell phenotype-dependent transcriptional regulation of
RT the alpha1 integrin gene.";
RL J. Biol. Chem. 272:26643-26651(1997).
DR EMBL; AB000470; BAA23160.1; -;
DR HSSP; AB000471; BAA23161.1; -;
DR IPR000413; Integrin_alpha.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF01839; FG-GAP; 5.
DR Pfam; PF00357; Integrin_A; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00191; Int_alpha; 4.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
DR PROSITE; PS50234; VWF_A; 1.
DR Integrin.
KW INTEGRIN.
SQ SEQUENCE 1171 AA; 130228 MW; B5054MD6F09736E CRC64;
Query Match 1.1%; Score 13; DB 13; Length 1171;
Best Local Similarity 100.0%; Pred. No. 0.0011;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 VIVLDGNSNTPW 181
DB 164 VIVLDGNSNTPW 176

RESULT 2

O29124 PRELIMINARY; PRT; 191 AA.
ID Q29124
AC Q29124
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DE 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=ENDOTHELIAL CELLS;
RX MEDLINE=95036279; PubMed=7949129;
RA Bahou W.F., Potter C.L., Mirza H.;
RT "The VLA-2 (alpha 2 beta 1) I domain functions as a ligand-specific
RT recognition sequence for endothelial cell attachment and spreading;
RT molecular and functional characterization.";
RL Blood 84:3734-3741(1994).
DR EMBL; Z12137; CAA78125.1; -;
DR HSSP; P17301; IAOX.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS50234; VWF_A; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 191 AA; 20896 MW; 2E228B472EC699D8 CRC64;

Query Match 0.9%; Score 11; DB 5; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 272 VVVTGESHG 282
DB 92 VVVTGESHG 102

RESULT 3

O9R7B6 PRELIMINARY; PRT; 383 AA.
ID O9R7B6
AC O9R7B6
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DE 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE Carbon starvation protein (Fragment).
GN CSG.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874;
RA Beddek A.J., O'Toole P.W.;
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U81796; AAD00522.1; -;
DR InterPro; IPR003706; CstA.
DR Pfam; PF02554; CstA; 1.
FT NON_TER 383
FT NON_TER 383
SQ SEQUENCE 383 AA; 41074 MW; C8F75BCBA1ADE9F CRC64;

Query Match 0.9%; Score 11; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1123 LMLIGSVLGG 1133
DB 118 LMLIGSVLGG 128

RESULT 4

O8WY18 PRELIMINARY; PRT; 823 AA.
ID O8WY18
AC O8WY18
DT 01-MAR-2002 (TRENBLREL. 20, Created)
DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
DE 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
OS MST018.
GN MST018.

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=AOA7A;
 RA Xu Y.Y., Sun L.Z., Wu Q.Y., Liu Y.Q., Liu B., Zhao B., Wang X.Y.,
 Song L., Ye J., Sheng H., Gao Y., Zhang C.L., Zhang J., Wei Y.J.,
 Sun Y.H., Jiang Y.X., Zhao X.W., Liu S., Liu L.S., Ding J.F.,
 Gao R.L., Qiang B.Q., Yuan J.G., Liew C.C., Zhao M.S., Hui R.T.;
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF111799; AAL39001.1; -
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam: PF01839; FG-GAP; 4.
 DR PRINTS: PR01185; INTEGRINA.
 DR SMART: SM00191; Int_alpha; 4.
 SQ SEQUENCE 823 AA; 92672 MW; DE4E78079DCD4925 CRC64;

Query Match 0.9%; Score 11; DB 4; Length 823;
 Best Local Similarity 100.0%; Pred. No. 0.09;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LCGLLALLLV 1141
 |||||
 DB 785 LCGLLALLLV 795

RESULT 5

Q06271 PRELIMINARY; PRT; 780 AA.
 ID Q06271;
 AC Q06271;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Integrin alpha-2 subunit (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Meng F., Desimone D.W.;
 RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 95-168 FROM N.A.
 RX MEDLINE=94008528; PubMed=8404528;
 RA Whitaker C.A., Desimone D.W.;
 RT "Integrin alpha subunit mRNAs are differentially expressed in early
 RT Xenopus embryos.";
 RL Development 117:1239-1249 (1993).
 DR EMBL: L43058; AAA69770.1; -
 DR EMBL: L10186; AAA16246.1; -
 DR HSP: P11215; IABX.
 DR InterPro: IPR000413; Integrin_alpha.
 DR Pfam: PF01839; FG-GAP; 4.
 DR PRINTS: PR01185; INTEGRINA.
 DR SMART: SM00191; Int_alpha; 4.
 DR PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 FT NON TER 1
 SQ SEQUENCE 780 AA; 87017 MW; 951B18C2B6BF637 CRC64;

Query Match 0.9%; Score 10; DB 13; Length 780;
 Best Local Similarity 100.0%; Pred. No. 0.92;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFGS 488
 |||||
 DB 86 GEOIGSYFGS 95

RESULT 6
 Q943Y3 PRELIMINARY; PRT; 66 AA.
 ID Q943Y3;
 AC Q943Y3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE P0552C05.18 protein.
 OS P0552C05.18.
 GN Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsunoto T., Yamamoto K.;
 RT "Oryza sativa nippohare (GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0552C05.";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AP002873; BAB64033.1; -
 SQ SEQUENCE 66 AA; 7136 MW; 196108BABC7972E2 CRC64;

Query Match 0.8%; Score 9; DB 10; Length 66;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 866 ECAAPSAAH 874
 |||||
 DB 11 ECAAPSAAH 19

RESULT 7
 Q922R3 PRELIMINARY; PRT; 84 AA.
 ID Q922R3;
 AC Q922R3;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Hypothetical protein RA0386.
 GN RA0386 OR SMA0726.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Rhizobium psyma (megaplasmid 1).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
 Gurley M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
 Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSyma megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
 DR EMBL: AE007230; AAK65044.1; -
 KW Hypothetical protein; plasmid; Complete proteome.
 SQ SEQUENCE 84 AA; 9792 MW; 5A847AB033610F9F CRC64;

Query Match 0.8%; Score 9; DB 16; Length 84;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 RLSPRRRL 729
 |||||
 DB 58 RLSPRRRL 66

RESULT 8

```

064016
ID 064016 PRELIMINARY; PRT; 236 AA.
AC 064016;
DT 01-ANG-1998 (TrEMBLrel. 07, Created)
DT 01-ANG-1998 (TrEMBLrel. 07, Last sequence update)
DE Putative lipoprotein.
DE Putative lipoprotein.
GN YOKB.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RP [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,
RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis SPBc2
RT prophage."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF020713; AAC12975.1; -.
KW Lipoprotein.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4F4AC CRC64;

Query Match 0.8%; Score 9; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164
DB 209 PEEKREK 217

RESULT 9
032005 PRELIMINARY; PRT; 236 AA.
AC 032005;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE YOKB protein.
DE YOKB protein.
GN YOKB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RP [1]
RP SEQUENCE FROM N.A.
RA STRAIN=168;
RA MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Betero M.G., Bessieres P., Bolochin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Comercon I.F., Cummings N.J., Daniel R.A.,
RA Denton F., Devine K.M., Duesterhoft A., Enrich S.D., Emmerson P.T.,
RA Eutlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Haelech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hull M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krog S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleier S., Schwoerer R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seior S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Teipstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

```

```

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99115; CAB14083.1; -.
KW Complete proteome.
SQ SEQUENCE 236 AA; 27337 MW; EAA0C26CDCF4F4AC CRC64;

Query Match 0.8%; Score 9; DB 16; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1156 PEEKREK 1164
DB 209 PEEKREK 217

RESULT 10
09RJL0 PRELIMINARY; PRT; 244 AA.
AC 09RJL0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Putative secreted protein.
DE Putative secreted protein.
GN SC00378 OR SCF62.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RP [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147 (2002).
DR EMBL; AL121855; CAB58319.1; -.
SQ SEQUENCE 244 AA; 25456 MW; 7C5575F67109D86B CRC64;

Query Match 0.8%; Score 9; DB 16; Length 244;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 565 ADVAVGAPL 573
DB 171 ADVAVGAPL 179

RESULT 11
08XSH5 PRELIMINARY; PRT; 399 AA.
AC 08XSH5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transport transmembrane protein.
GN RSP0499 OR RS00378.

```


OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681679; PubMed=11823852;
 RA Salanoubat M., Genin S., Arciguenave F., Gouzy J., Mangenot S.,
 RA Arlaud M., Billault A., Brotier P., Camus J.C., Cattoilco L.,
 RA Chardin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siglier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502 (2002).
 DR EMBL; AL646079; CAD17650.1; -;
 DR InterPro; IPR001179; FKBP_PPIase.
 DR PROSITE; PS00453; FKBP_PPIase_1; UNKNOWN_1.
 KW Plasmid; Complete proteome.
 SQ SEQUENCE 399 AA; 42148 MW; 096BB97E83DCA7D1 CRC64;

Query Match 0.8%; Score 9; DB 16; Length 399;
 Best Local Similarity 100.0%; Pred. No. 5.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLTLALV 1141
 DB 161 GLLTLALV 169

RESULT 12

O94LU9 PRELIMINARY; PRT; 757 AA.

AC O94LU9; 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Putative
 DE UDP-N-acetyl[muramoyl]anaryl-D-glutamate-2,6-diaminopimelate
 DE ligase.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacridae; Oryzaeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Tsirlin T.,
 RA Riggs P., Hsiao J., Zismann T.V., Blunt S., Pai G., Vanaken S.E.,
 RA Uterback T.R., Feldlyum T.V., Quackenbush J., Salzberg S.L.,
 RA White O., Frazer C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa0082M15 genomic sequence";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC020666; AK43503.1; -;
 DR InterPro; IPR000713; Mur_Ligase.
 DR InterPro; IPR004101; Mur_Ligase_C.
 DR Pfam; PF01225; Mur_Ligase; 1.
 DR Pfam; PF02875; Mur_Ligase_C; 1.
 SQ SEQUENCE 757 AA; 82194 MW; 8EF030AFAD51F80 CRC64;

Query Match 0.8%; Score 9; DB 10; Length 757;
 Best Local Similarity 100.0%; Pred. No. 9.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 VAVGAPLE 575
 DB 543 VAVGAPLE 551

RESULT 13
 O96HB1 PRELIMINARY; PRT; 1086 AA.
 AC O96HB1; 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Unknown (protein for MGC:1714).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RA Strussberg R.;
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC008777; AAH08777.1; -;
 DR InterPro; IPR00413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; UNKNOWN_1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1086 AA; 119223 MW; F6FF2546B8C32F9 CRC64;

Query Match 0.8%; Score 9; DB 4; Length 1086;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574
 DB 455 DVAVGAPLE 463

RESULT 14

O9R200 PRELIMINARY; PRT; 1160 AA.

AC O9R200; 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Integrin alpha L.
 OS ITGAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SPLEEN;
 RA Ma R.Z., Teuscher C.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF065901; AAD25884.1; -;
 DR HSSP; P20701; ILPA.
 DR MGD; MGI:96606; Itgal.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PR01185; INTEGRINA.
 DR PRINTS; PR00453; VWFADOMAIN.
 DR SMART; SM00191; Itc_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1160 AA; 128127 MW; A33C531B139F1PAD CRC64;

Query Match 0.8%; Score 9; DB 11; Length 1160;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 |||||
 DB 536 DVAVGAPLE 544

RESULT 15

Q9WTV4 PRELIMINARY; PRT; 1161 AA.

DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE Integrin alpha L.
 GN ITGAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBA/2J; TISSUE=SPLEEN;
 RA Ma R.Z., Teuscher C.,
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 DR HSSP; P20701; ILFA.
 DR MGD; MGJ:96606; Itgal.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00242; INTEGRIN_ALPHA; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1161 AA; 128240 MM; 86B102F7B209E431 CRC64;

Query Match 0.8%; Score 9; DB 11; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 |||||
 DB 536 DVAVGAPLE 544

RESULT 16

Q98TF0 PRELIMINARY; PRT; 1187 AA.

AC Q98TF0;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE CD11-2.
 GN CIA2.
 OS Cyprinus carpio (Common carp).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprinus.
 OX NCBI_TaxID=7962;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=PERITONEAL EXUDATE CELLS;
 RA Kimura M., Fujiki K., Nakao M.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB048537; BAB39135.1; -.
 DR HSSP; P20701; ILFA.
 DR InterPro; IPR001969; Asprotease_site.
 DR InterPro; IPR001969; Asprotease_site.
 DR InterPro; IPR000413; Integrin_alpha.

DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE; PS50234; VWF_A; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1187 AA; 131778 MM; 85EDC7CA8B6B1C59 CRC64;

Query Match 0.8%; Score 9; DB 13; Length 1187;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFG 487
 |||||
 DB 454 GEOIGSYFG 462

RESULT 17

Q98TF1 PRELIMINARY; PRT; 1196 AA.

AC Q98TF1;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
 DE CD11-1.
 GN CIA1.
 OS Cyprinus carpio (Common carp).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprinus.
 OX NCBI_TaxID=7962;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=PERITONEAL EXUDATE CELLS;
 RA Kimura M., Nakao M., Mura C., Fujiki K., Yano T.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB048536; BAB39134.1; -.
 DR HSSP; P20701; ILFA.
 DR InterPro; IPR001969; Asprotease_site.
 DR InterPro; IPR000413; Integrin_alpha.
 DR InterPro; IPR002035; VWF_A.
 DR Pfam; PF01839; FG-GAP; 5.
 DR Pfam; PF00357; Integrin_A; 1.
 DR Pfam; PF00092; vwa; 1.
 DR PRINTS; PRO1185; INTEGRINA.
 DR PRINTS; PRO0453; VWFADOMAIN.
 DR SMART; SM00191; Int_alpha; 5.
 DR SMART; SM00327; VWA; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_1.
 DR PROSITE; PS50234; VWF_A; 1.
 DR PROSITE; PS50234; VWF_A; 1.
 SQ SEQUENCE 1196 AA; 132477 MM; 9369C807E7DCA53B CRC64;

Query Match 0.8%; Score 9; DB 13; Length 1196;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 479 GEOIGSYFG 487
 |||||
 DB 454 GEOIGSYFG 462

RESULT 18

Q9JUR7 PRELIMINARY; PRT; 59 AA.

AC Q9JUR7;
 DT 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last annotation update)

```

DE LST-1/K protein.
GN LST-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BLOOD;
RA Rollinger-Holinger I., Eibl B., Pauly M., Weiss E.;
RT "LST1 gene with complex alternative splicing and immunosuppressive
RT function." (DEC-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y18487; CAB59905.1; -.
SQ SEQUENCE 59 AA; 6122 MW; FC2FF156D576C486 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 59;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1131 LGGLLLA 1138
DB 16 LGGLLLA 23

RESULT 19
O9MC41 PRELIMINARY; PRT; 77 AA.
ID O9MC41.
AC O9MC41.
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Oref1.
GN Oref1.
OS Bacteriophage D3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC Lambda-like viruses.
OX NCBI_TaxID=31535;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20042341; PubMed=10572124;
RA Gilakjan Z.A., Kropinski A.M.;
RT "Cloning and analysis of the capsid morphogenesis genes of Pseudomonas
RT aeruginosa bacteriophage D3; another example of protein chain malty";
RL J. Bacteriol. 181:7221-7227(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20485557; PubMed=11029426;
RA Kropinski A.M.;
RT "Sequence of the Genome of the Temperate, Serotype-Converting,
RT Pseudomonas aeruginosa Bacteriophage D3.";
RL J. Bacteriol. 183:6066-6074(2000).
DR EMBL; AF165214; AAF80767.1; -.
SQ SEQUENCE 77 AA; 8266 MW; 39DAB91077F7234B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 9; Length 77;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 756 TTTTALDN 763
DB 42 TTTTALDN 49

RESULT 20
O00453 PRELIMINARY; PRT; 97 AA.
ID O00453.
AC O00453.
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE LST1 protein.
GN LST1.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98035883; PubMed=9367684;
RA de Baey A., Fellerhoff B., Mater S., Martinuzzi S., Weidie U.,
RA Weiss E.H.;
RT "Complex expression pattern of the TNF region gene LST1 through
RT differential regulation, initiation, and alternative splicing.";
RL Genomics 45:581-600(1997).
DR EMBL; AF000425; AAB86999.1; -.
SQ SEQUENCE 97 AA; 10822 MW; AA03C761E787AF94 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 97;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1131 LGGLLLA 1138
DB 16 LGGLLLA 23

RESULT 21
O8XEP2 PRELIMINARY; PRT; 103 AA.
ID O8XEP2.
AC O8XEP2.
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Putative septum formation initiator (Hypothetical protein
DE STY3056).
GN YGBQ OR STM2931 OR STY3056.
OS Salmonella typhimurium, and
OC Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Malyan E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
RA Baker S., Basham D., Brooks R., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AE008833; AAL2181.1; -.
DR EMBL; AL627276; CAD06037.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11575 MW; F3B55AB77E8AC6C CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 103;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1134 LLLALLV 1141

Db 6 LLLALLV 13

RESULT 22

Q9EY3

ID Q9EY3 PRELIMINARY; PRT; 105 AA.

AC Q9EY3; 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, last sequence update)

DE 01-MAR-2001 (TReMBLrel. 16, last annotation update)

OS Ygbo.

OC Klebsiella aerogenes.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OX NCBI_TaxID=28451;

RN (1) SEQUENCE FROM N.A.

RC STRAIN=W70; MEDLINE=20566700; PubMed=11114933;

RA Kolko M.M., Kapetanovich L.A., Lawrence J.G.;

RT "Alternative pathways for siroheme synthesis in Klebsiella aerogenes."

RL J. Bacteriol. 183:328-335(2001).

RN (2) SEQUENCE FROM N.A.

RC STRAIN=W70; Seifeldin T.A., Lawrence J.G.;

RA "Methionine recycling in Klebsiella aerogenes."

RT J. Bacteriol. 0:0-0(2001).

DR EMBL; AF308468; AAG42461.1; -

SQ SEQUENCE 105 AA; 11951 MW; 530471363FD3112A CRC64;

Query Match 0.7%; Score 8; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1134 LLLALLV 1141

Db 6 LLLALLV 13

RESULT 23

Q8RIE9

ID Q8RIE9 PRELIMINARY; PRT; 112 AA.

AC Q8RIE9; 01-JUN-2002 (TReMBLrel. 21, Created)

DT 01-JUN-2002 (TReMBLrel. 21, last sequence update)

DE 01-JUN-2002 (TReMBLrel. 21, last annotation update)

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN (1) SEQUENCE FROM N.A.

RC TISSUE=SALIVARY GLAND;

RA Strausberg R.;

RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC024677; AAH24677.1; -

KW Hypothetical protein.

SQ SEQUENCE 112 AA; 12371 MW; 69CE957825183F72 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 112;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1134 LLLALLV 1141

Db 6 LLLALLV 13

Db 5 LLLALLV 12

RESULT 24

Q9OX14

ID Q9OX14 PRELIMINARY; PRT; 119 AA.

AC Q9OX14; 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)

DE 01-MAY-2000 (TReMBLrel. 13, last annotation update)

OS Rat.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN (1) SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;

RA Ragnunathan A., Sivakamasundari R., Weisman S.M.;

RT "B14 homolog expressed in Rat."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF208230; AAF20145.1; -

SQ SEQUENCE 119 AA; 13382 MW; B7E5EB8B874E3583 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 119;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 VIGGLLL 1137

Db 41 VIGGLLL 48

RESULT 25

Q9RSB5

ID Q9RSB5 PRELIMINARY; PRT; 127 AA.

AC Q9RSB5; 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)

DE 01-MAR-2002 (TReMBLrel. 20, last annotation update)

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;

OC Deinococcaceae; Deinococcus.

OX NCBI_TaxID=1299;

RN (1) SEQUENCE FROM N.A.

RC STRAIN=R1;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.;

RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.;

RA Vamathevan J.J., Lam P., McDonald L., Utechtack T., Zaleski C.;

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.;

RA Fraser C.M.;

RT "Genome sequence of the radioresistant bacterium Deinococcus

RT radiodurans R1."

RT Science 286:1571-1577(1999).

DR EMBL; AE002054; AAF11763.1; -

KW Hypothetical protein. Complete proteome.

SQ SEQUENCE 127 AA; 14314 MW; 2B3E679D3522A0B1 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 127;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1133 GLLALL 1140

Db 7 GLLALL 14

#

RESULT 26

O9KNZ1

PRELIMINARY;

PRT; 136 AA.

AC O9KNZ1;

01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Ribosomal protein L16.

GN VCS589.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OC NCBI_TaxID=666;

[1]

RC SEQUENCE FROM N.A.

STRAIN=EL TOR N1691 / SEROTYPE O1;

MEDLINE=20406833; PubMed=10952201;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unyam L.A.,

Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

Ermoiaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,

McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,

Salberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

Frazer C.M.;

"DNA sequence of both chromosomes of the cholera pathogen Vibrio

cholerae.";

RT Nature 406:477-483 (2000).

RL EMBL; AE004326; AAF95730.1; -.

DR TIGR; VC2589; -.

DR InterPro; IPR000114; Ribosomal_L16.

DR Pfam; PF00252; Ribosomal_L16; 1.

DR PRINTS; PR00060; RIBOSOMALL16.

DR TIGRFAMs; TIGR01164; rplp_bact; 1.

DR PROSITE; PS00586; RIBOSOMAL_L16_1; 1.

KM Complete proteome.

SQ SEQUENCE 136 AA; 15551 MW; 3F74B546B61F917D CRC64;

Query Match

Best Local Similarity 0.7%; Score 8; DB 16; Length 136;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1055 LAKGTEVS 1062

DB 20 LAKGTEVS 27

RESULT 27

O61692

PRELIMINARY;

PRT; 141 AA.

AC O61692;

01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE Putative heat stable antigen.

GN HSA-C.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

[1]

RN SEQUENCE FROM N.A.

STRAIN=CBA X C57BL/6; TISSUE=SPLEEN;

RC MEDLINE=91209380; PubMed=2019286;

RA Wenger R.H., Ayane M., Bose R., Koehler G., Nielsen P.J.;

"The genes for a mouse hematopoietic differentiation marker called the

heat-stable antigen.";

RT Eur. J. Immunol. 21:1039-1046 (1991).

RL EMBL; X56486; CAA39843.1; -.

DR TIGRFAMs; TIGR00079; pepc_deformylase; 1.

KM Complete proteome.

SQ SEQUENCE 141 AA; 15515 MW; E4BFB428ADC03C69 CRC64;

Query Match

Best Local Similarity 0.7%; Score 8; DB 11; Length 141;

Matches 100.0%; Pred. No. 24;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLLLALL 1140

DB 12 GLLLLALL 19

RESULT 28

O8CG30

PRELIMINARY;

PRT; 146 AA.

AC O8CG30;

01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Bradykinin-potentiating protein.

OS Bothrops insularis (Island jararaca) (Quelana jararaca).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidodonta; Squamata; Serpentes; Colubridae;

OC Viperae; Crotalinae; Bothrops.

OC NCBI_TaxID=8723;

[1]

RN SEQUENCE FROM N.A.

TISSUE=VENOM GLAND;

Junqueira-de-Azevedo I.L.M., Ho P.L.;

"A survey of gene expression and diversity in the venom glands of the

pitviper Bothrops insularis through the generation of Expressed

Sequence Tags (ESTs).";

RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL; AF490532; AAM09691.1; -.

DR EMBL; AF490532; AAM09691.1; -.

SQ SEQUENCE 146 AA; 16088 MW; F58244C4F24B1799 CRC64;

Query Match

Best Local Similarity 0.7%; Score 8; DB 13; Length 146;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLLLALL 1140

DB 10 GLLLLALL 17

RESULT 29

O8ZLM7

PRELIMINARY;

PRT; 169 AA.

AC O8ZLM7;

01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Peptide deformylase (EC 3.5.1.31).

GN DEF OR STM3406.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC OC Salmonella.

OC NCBI_TaxID=602;

[1]

RN SEQUENCE FROM N.A.

STRAIN=LT2 / SGSC1412 / ATCC 700720;

RC MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,

Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

Waterston R., Wilson R.K.;

"Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

RT Nature 413:852-856 (2001).

RL EMBL; AE008857; AAL22269.1; -.

DR InterPro; IPR000181; pep_deformylase.

DR Pfam; PF01327; pep_deformylase; 1.

DR PRINTS; PR01576; PDEFORMYLASE.

DR Prodom; PD003844; pep_deformylase; 1.

DR TIGRFAMs; TIGR00079; pepc_deformylase; 1.

KM Hydrolase; Complete proteome.

SQ SEQUENCE 169 AA; 19282 MW; 3181AB2F6BF02765 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 169;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198
 |||||
 DB 138 LVGKLFID 145

RESULT 30

Q821W9 PRELIMINARY; PRT; 169 AA.
 AC Q821W9;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Polypeptide deformylase (EC 3.5.1.31).
 GN FMS OR STY4391.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella
 NCBI_TaxID=601;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; Pubmed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 Cronin A., Davis P., Davies R.M., Dowd L., Farrar J.,
 Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsis K.,
 Krogh A., Larsen T.S., Leather S., Moulie S., O'Goora P., Parry C.,
 Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 Whitehead S., Barrett B.G.,
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627282; CAD09179.1; -;
 DR InterPro: IPR000181; Pep.deformylase.
 DR Pfam: PF01327; Pep.deformylase; 1.
 DR PRINTS: PR01576; PDEFORMLAS.
 DR ProDom: PD003844; Pep.deformylase; 1.
 DR TrEMBL: TIGR00079; pepc.deformyl; 1.
 KW Hydroxylase; Complete proteome.

SO SEQUENCE 169 AA; 19281 MW; 618BA12F6BF02ADE CRC64;

Query Match 0.7%; Score 8; DB 16; Length 169;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 LVGKLFID 198
 |||||
 DB 138 LVGKLFID 145

RESULT 31

Q90Y12 PRELIMINARY; PRT; 181 AA.
 AC Q90Y12;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Bradykinin potentiating peptide and C-type natriuretic peptide isoform
 DE 1.
 OS *Crotalus durissus terrificus* (South American rattlesnake).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidodermata; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Viperidae; Crotalinae; Crotalus.
 NCBI_TaxID=8732;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=VENOM GLAND;

RA Hayashi M.A.F., Radis-Baptista G., Barbosa S.R., Yamane T.,
 RA Camargo A.C.M.;
 RT "Crotalus durissus terrificus bradykinin potentiating peptide
 RT precursor.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF308593; AAL09426.1; -;
 DR InterPro: IPR000663; Natr_peptide.
 DR Pfam: PF00212; ANP; 1.
 DR PROSITE: PS00263; NATRIURETIC PEPTIDE; UNKNOWN 1.
 SO SEQUENCE 181 AA; 18560 MW; 7B5ADC5B9372D07F CRC64;

QY 1133 GLLTALL 1140
 |||||
 DB 10 GLLTALL 17

Query Match 0.7%; Score 8; DB 13; Length 181;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 32

Q90Y11 PRELIMINARY; PRT; 181 AA.
 AC Q90Y11;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Bradykinin potentiating peptide and C-type natriuretic peptide isoform
 DE 2.
 OS *Crotalus durissus terrificus* (South American rattlesnake).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidodermata; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Viperidae; Crotalinae; Crotalus.
 NCBI_TaxID=8732;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=VENOM GLAND;
 RA Hayashi M.A.F., Radis-Baptista G., Barbosa S.R., Yamane T.,
 RA Camargo A.C.M.;
 RT "Crotalus durissus terrificus bradykinin-potentiating peptide and C-
 RT type natriuretic peptide precursor isoform2.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF308594; AAL09427.1; -;
 DR InterPro: IPR000663; Natr_peptide.
 DR Pfam: PF00212; ANP; 1.
 DR PROSITE: PS00263; NATRIURETIC PEPTIDE; UNKNOWN 1.
 SO SEQUENCE 181 AA; 18507 MW; 9B2E95D38A5FF27 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 181;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLTALL 1140
 |||||
 DB 10 GLLTALL 17

RESULT 33

Q97LC8 PRELIMINARY; PRT; 185 AA.
 AC Q97LC8;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Predicted membrane protein.
 GN CAC0634.
 OS *Clostridium acetobutylicum*.
 OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 NCBI_TaxID=1488;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

```

RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.D.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007579; AAK78611.1; -.
KM Complete proteome.
SQ SEQUENCE 185 AA; 20003 MW; DEE2B40DD86612CB CRC64;

Query Match 0.7%; Score 8; DB 16; Length 185;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 636 AALLSSR 643
DB 54 AALLSSR 61

RESULT 34
Q9YBV7 PRELIMINARY; PRT; 240 AA.
AC Q9YBV7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Hypothetical protein APL1493.
GN APL1493.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaei A., Koeugl H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kusuda N., Oguchi A., Aoki K.-I., Kudota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000061; BAA80491.1; -.
DR InterPro; IPR003754; HEM4.
DR Pfam; PF02602; HEM4; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 240 AA; 25442 MW; 14EB211AD0646A7A CRC64;

Query Match 0.7%; Score 8; DB 17; Length 240;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 432 GYSVSSML 439
DB 107 GYSVSSML 114

RESULT 35
Q9RK95 PRELIMINARY; PRT; 258 AA.
AC Q9RK95;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative hydrolase.
GN SC00267 OR SCF1.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL117322; CAB55529.1; -.
DR InterPro; IPR000073; Abhydrolase.
DR InterPro; IPR003089; AB hydrolase.
DR InterPro; IPR000379; Ser esterase site.
DR Pfam; PF00561; abhydrolase; 1.
DR PRINTS; PR00111; ABHYDROLASE.
KM Hydrolase.
SQ SEQUENCE 258 AA; 27454 MW; 4E9B7CFECA5802C5 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 258;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 284 ELPAALKA 291
DB 156 ELPAALKA 163

RESULT 36
Q9PW56 PRELIMINARY; PRT; 265 AA.
AC Q9PW56;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Bradykinin-potentiating/C-type natriuretic peptide.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Hayaishi M.A.F., Murbach A.F., Camargo A.C.M.;
RT "The precursor of C-type natriuretic peptide of snake brain contains
RT angiotensin converting enzyme inhibitors, specific for the C-catalytic

```

RT site."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF171670; AAD51326.2; -
 DR InterPro: IPR000663; Nact_peptide.
 DR Pfam: PF00212; ANP; 2
 DR PRINTS: PRO0710; NATPEPTIDES.
 DR SMART: SM00183; NAT_PEP; 2.
 DR PROSITE: PS00263; NATRIURETIC PEPTIDE; 2.
 SQ SEQUENCE 265 AA; 27763 MW; 8E99AEC976CCD439 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 265;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLLALL 1140
 |||||
 10 GLLLLALL 17

RESULT 37
 08Q991 PRELIMINARY; PRT; 265 AA.
 AC 08Q991;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Bradykinin-potentiating/c-type natriuretic protein.
 OS Bothrops insularis (Island jararaca) (Queimada jararaca).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidae;
 OC Viperidae; Crotalinae; Bothrops.
 OX NCBI_TaxID=8723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=VENOM GLAND;
 RC Unpublished-Azevedo I.L.M., Ho P.L.;
 RA "A survey of gene expression and diversity in the venom glands of the
 RT pitviper Bothrops insularis through the generation of Expressed
 RT Sequence Tags (ESTs).";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF490531; AAM09690.1; -
 SQ SEQUENCE 265 AA; 27763 MW; 0EAE1408B42358BE CRC64;

Query Match 0.7%; Score 8; DB 13; Length 265;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1133 GLLLLALL 1140
 |||||
 10 GLLLLALL 17

RESULT 38
 09U168 PRELIMINARY; PRT; 269 AA.
 ID 09U168;
 AC 09U168;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE Hypothetical 28.8 kDa protein.
 GN L377.08.
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OC NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;
 RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Murphy L., Quail M.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;

RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL: AL136326; CAB65934.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 269 AA; 28805 MW; F35F62D65882141A CRC64;

Query Match 0.7%; Score 8; DB 5; Length 269;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 600 AAASMPHA 607
 |||||
 75 AAASMPHA 82

RESULT 39
 053905 PRELIMINARY; PRT; 281 AA.
 AC 053905;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Actva 3 protein.
 GN ACTVA 3.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE=92114870; PubMed=1766437;
 RA Caballero J.L., Martinez E., Malpartida F., Hopwood D.A.;
 RA "Organization and functions of the actva region of the actinorhodin
 RT biosynthetic gene cluster of Streptomyces coelicolor.";
 RL Mol. Gen. Genet. 230:401-412(1991).
 DR EMBL: X58833; CAA41639.1; -
 SQ SEQUENCE 281 AA; 29952 MW; 55EC061CE3C2EFB3 CRC64;

Query Match 0.7%; Score 8; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 708 TAGARAAF 715
 |||||
 89 TAGARAAF 96

RESULT 40
 0931Z2 PRELIMINARY; PRT; 281 AA.
 ID 0931Z2;
 AC 0931Z2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical protein actva3.
 GN ACTVA3 OR SCOS078 OR SCBAC28G1.04.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kleser H., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 DR EMBL; A0593842; CAC44191.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 281 AA; 30028 MW; DFC4A496C340E1B3 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB:16; Length 281;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 708 TAGARAAF 715
 |||||
 DB 89 TAGARAAF 96

RESULT 41
 ID 0930X2 PRELIMINARY; PRT; 287 AA.
 AC 0930X2;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Lycopen elongase.
 GN CRTB.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
 OC Actinomycetales; Corynebacteriineae; Corynebacteriaceae;
 OC Corynebacterium.
 OC NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=MU233;
 RA Krubasik P., Kobayashi M., Sandmann G.;
 RT "Expression and functional analysis of a gene cluster involved in the
 RT synthesis of decaprenoxanthin reveals the mechanisms for CSO
 RT carotenoid formation";
 RL Eur. J. Biochem. 286:3703-3708(2001).
 DR EMBL; AF159510; AKK64302.1; -;
 DR InterPro; IPR000537; UblA.
 DR Pfam; PF01040; UblA; 1.
 SQ SEQUENCE 287 AA; 31674 MW; 3FD29E1CDDASFOA CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB:2; Length 287;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 ILUSRP1 645
 |||||
 DB 8 ILUSRP1 15

RESULT 42
 ID 09XCM9 PRELIMINARY; PRT; 294 AA.
 AC 09XCM9;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE NoSF.
 GN NOSF.
 OS Rhodobacter sphaeroides f. sp. denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OC NCBI_TaxID=39723;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL106;
 RX MEDLINE=99429850; PubMed=10498715;
 RA Sabaty M., Schwintner C., Cahors S., Richard P., Vermeglio A.;

RT "Nitrite and nitrous oxide reductase regulation by nitrogen oxides in
 RT Rhodobacter sphaeroides f. sp. denitrificans Il106.";
 RL J. Bacteriol. 181:6028-6032(1999).
 DR EMBL; AF15260; AAD3475.1; -;
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR003439; ABC_transport.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transport; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
 DR ATP-binding.
 KW ATP-binding.
 SQ SEQUENCE 294 AA; 31061 MW; C76B100453535EF5 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB:2; Length 294;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 GAAILLS 642
 |||||
 DB 176 GAAILLS 183

RESULT 43
 ID 084105 PRELIMINARY; PRT; 298 AA.
 AC 084105;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE HAD superfamily hydrolase/phosphatase.
 GN CT103.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=D/UM-3/CX;
 RA MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 DR EMBL; AE001284; AAC67694.1; -;
 DR Hydrolase; Complete proteome.
 KW Hydrolase; Complete proteome.
 SQ SEQUENCE 298 AA; 34254 MW; 9F2ADCCFDF67A74D CRC64;

Query Match
 Best Local Similarity 100.0%; Score 8; DB:16; Length 298;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 265 PEARLIV 272
 |||||
 DB 141 PEARLIV 148

RESULT 44
 ID 09R9K7 PRELIMINARY; PRT; 299 AA.
 AC 09R9K7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE NoSF protein.
 GN NOSF.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Paracoccus.
 OC NCBI_TaxID=266;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PD1222;

```

RX MEDLINE=20416234; PubMed=10960107;
RA Saudere N.F.W., Hrnberg J.J., Reijnders W.N.M., Westerhoff H.V.,
RA de Vries S., van Spanning R.J.M.;
RT "The NosX and Nix proteins of Paracoccus denitrificans are functional
RT homologues: Their role in maturation of nitrous oxide reductase.";
RL J. Bacteriol. 182:5211-5217(2000).
DR EMBL; A010260; CAB5353.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00005; ABC_tran.1.
DR ProDom; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KM ATP-binding.
SQ SEQUENCE 299 AA; 31687 MW; 69849D3F85456214 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 299;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 635 GAATLSS 642
DB 179 GAATLSS 186

RESULT 45
Q94K03 PRELIMINARY; PRT; 299 AA.
AC Q94K03;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Unknown protein.
GN A_IG002N01.8
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK370497; AAK43874.1; -.
SQ SEQUENCE 299 AA; 34028 MW; BB2AB03CA4184D3E CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 299;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
DB 44 LLLALLV 51

RESULT 46
O05656 PRELIMINARY; PRT; 309 AA.
AC O05656;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative malate oxidoreductase (Fragment).
GN MAOX.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10987;
RA Koltoe A.B.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10987;
RX MEDLINE=99231848; PubMed=10217496;
RA Oksrad O.A., Hegna I., Lindback T., Rishovd A.L., Kolsto A.B.;
RT "Genome organisation is not conserved between Bacillus cereus and
RT Bacillus subtilis.";
RL Microbiology 145:621-631(1999).
DR EMBL; Y09212; CAA70412.1; -.
DR InterPro; IPR001891; Malic_oxred.
DR Pfam; PF00390; malic; 1.
DR PRINTS; PR00072; MALOXRDASE.
DR PROSITE; PS00331; MALIC_ENZYMES; 1.
FT NON_TER 1
FT NON_TER 309
SQ SEQUENCE 309 AA; 34451 MW; 5FF0C7A9BF0470EF CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 309;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 271 LVVVTDE 278
DB 170 LVVVTDE 177

RESULT 47
O8TES5 PRELIMINARY; PRT; 348 AA.
AC O8TES5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE FLJ00114 protein (Fragment).
GN FLJ00114.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT "The nucleotide sequence of a long CDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074047; BAB84873.1; -.
FT NON_TER 1
FT NON_TER 348
SQ SEQUENCE 348 AA; 38941 MW; 5A698D66A1C9992 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 348;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 300 GGLLLAL 307

RESULT 48
Q9M146 PRELIMINARY; PRT; 360 AA.
AC Q9M146;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Hypothetical 41.1 kDa protein.
GN AT4G01220.

```

OC Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL161491; CAB80931.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 360 AA; 41120 MW; A7E4CBAB9369D931 CRC64;
 Query Match 0.7%; Score 8; DB 10; Length 360;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1134 LLLALLV 1141
 Db 44 LLLALLV 51
 RESULT 49
 016339
 ID 016339 PRELIMINARY; PRT; 383 AA.
 AC 016339;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE F59P6.2 protein.
 GN F59P6.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OK NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier W., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL N2;
 RA Bradshaw H., Graves T.;
 RL "The sequence of C. elegans cosmid F59D6.";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA STRAIN=BRISTOL N2;
 RC Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF016435; AAB65877.1; -
 DR HSSP; P00794; 4CMS.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 SQ SEQUENCE 383 AA; 42004 MW; D9DE356D3AB7401E CRC64;

Query Match 0.7%; Score 8; DB 5; Length 383;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1136 LLLALLVFC 1143
 Db 8 LLLALLVFC 15
 RESULT 50
 029F91
 ID 029F91 PRELIMINARY; PRT; 385 AA.
 AC 029F91;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Possible acyltransferase, copy 1.
 GN L8032.06.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OK NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
 RA Rajandream M.A., Barrell B.G.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIEDLIN;
 RX MEDLINE=96146435; PubMed=9477241;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL; AL139794; CAC22648.1; -
 DR InterPro; IPR002123; Acyltransferase.
 DR Pfam; PF01553; Acyltransferase; 1.
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 385 AA; 43595 MW; D4BBD5E7E8B9791 CRC64;
 Query Match 0.7%; Score 8; DB 5; Length 385;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 716 DSGGRLS 723
 Db 162 DSGGRLS 169
 RESULT 51
 024416
 ID 024416 PRELIMINARY; PRT; 405 AA.
 AC 024416;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Pectate lyase.
 GN PL.
 OS Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustids II; Rosales; Rosaceae; Rosoideae; Fragaria.
 OK NCBI_TaxID=3747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=CHANDLER;
 RX MEDLINE=97435972; PubMed=9290639;
 RA Medina-Escobar N., Cardenas J., Moyano E., Caballero J.L.,
 RA Munoz-Bianco J.;
 RT "Cloning, molecular characterization and expression pattern of a
 strawberry ripening-specific cDNA with sequence homology to pectate

RT lyase from higher plants.";
 RL Plant Mol. Biol. 34:867-877(1997).
 DR EMBL: U63550; AAB71208.1; -.
 DR InterPro: IPR002022; Amb_allergen.
 DR Pfam: PF00544; pec_lyase; 1.
 KW Lyase.
 SQ SEQUENCE 405 AA; 45744 MW; 563BC01BCA640559 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 405;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 18 LLLALLV 25

RESULT 52
 ID 094FT6 PRELIMINARY; PRT; 405 AA.
 AC 094FT6;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Pectate lyase B (Fragment).
 OC Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid 1; Rosales; Rosaceae; Rosoideae; Fragaria.
 OC NCBI_TaxID=3747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHANDLER;
 RA Munoz-Bianco J., Caballero J.L., Benitez-Burraco A.;
 RT Ripening-related strawberry pectate-lyase genes.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF339024; AAK66160.1; -.
 DR InterPro: IPR002022; Amb_allergen.
 DR Pfam: PF00544; pec_lyase; 1.
 KW Lyase.
 FT NON_TER 405 405.
 SQ SEQUENCE 405 AA; 45654 MW; DB1692E59BF8300 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 405;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 18 LLLALLV 25

RESULT 53
 ID 0915T3 PRELIMINARY; PRT; 415 AA.
 AC 0915T3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
 DE Probable permease of ABC transporter.
 GN PA0605.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 DR EMBL: AE004497; AAG03994.1; -.
 DR InterPro: IPR000515; BPD_transp.
 DR Pfam: PF00528; BPD_transp; 1.
 DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 415 AA; 45925 MW; A0F3425B8AE60468 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 415;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLA 1138
 DB 388 LGGLLLA 395

RESULT 54
 ID 08YGR5 PRELIMINARY; PRT; 433 AA.
 AC 08YGR5;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein Alr3754.
 GN ALR3754.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OC NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL: AP003594; BAB75453.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 433 AA; 48889 MW; 5991F41A9C80C0B4 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 433;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 982 VSGLIISA 989
 DB 247 VSGLIISA 254

RESULT 55
 ID 004611 PRELIMINARY; PRT; 444 AA.
 AC 004611;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE A_IG002N01.8 protein.
 GN A_IG002N01.8.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid 1; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=CV. COLUMBIA;
RA Scheet P., Maggi L.;
RT "The sequence of A. thaliana IG002N01.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Wash-U;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Waterston R.;
RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF007269; AAB61018.1; -.
SQ SEQUENCE 444 AA; 50340 MW; A670696CA0052DF4 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 444;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
DB 115 LLLALLLV 122

RESULT 56
Q8XZT4 PRELIMINARY; PRT; 473 AA.
ID Q8XZT4;
AC Q8XZT4;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Probable replicative DNA helicase protein (EC 3.6.1.-).
GN DNAB OR RSC1311 OR RS02833.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMT1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brotier P., Camus J.C., Cattolico L.,
RA Chandel M., Choise N., Claudel-Renaud C., Cunnac S., Demange N.,
RA Gaspin C., Lavie W., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siglier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
DR EMBL; AL646063; CAD15013.1; -.
DR InterPro; IPR003593; AAA_Arpase.
DR InterPro; IPR001198; DNAB_helicase.
DR Pfam; PF00772; DNAB; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00665; DNAB; 1.
DR HydroLase; Complete proteome.
SQ SEQUENCE 473 AA; 52293 MW; 155F5FC98B5434E6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 473;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1129 SVLGGILL 1136
DB 34 SVLGGILL 41

RESULT 57
Q9REL8 PRELIMINARY; PRT; 491 AA.

```

```

AC Q9REL8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Response regulator HupR.
GN HUPR.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=601;
RA Colbeau A., Chabert J., Wu J.-Q., Vignais P.M.;
RT "Hup gene cluster from Rhodobacter sphaeroides.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING
DOMAIN.
CC EMBL; AF214145; AAF19999.2; -.
DR HSSP; P41789; INTR.
DR InterPro; IPR002197; HTH_Fis.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR002078; Sig54_interact.
DR Pfam; PF02954; HTH_8; 1.
DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00158; Sigma54_activat; 1.
DR PRINTS; PR01590; HTHPIS.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00448; REC; 1.
DR TIGRFAMs; TIGR01199; HTH_fis; 1.
DR PROSITE; PS00688; SIGMA54_INTERACT_3; UNKNOWN_1.
DR PROSITE; PS00445; SIGMA54_INTERACT_4; 1.
KW ATP-binding; DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription regulation.
SQ SEQUENCE 491 AA; 53928 MW; 09D348AE76A328 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 491;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 RMALEDEF 420
DB 20 RMALEDEF 27

RESULT 58
Q9K559 PRELIMINARY; PRT; 491 AA.
ID Q9K559;
AC Q9K559;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative hydogenase transcriptional regulator (HupR protein).
GN HUPR.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=601;
RA Xu D., Liang J., Wu Y.;
RT "The study of the hupR gene in Rhodobacter sphaeroides 601.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=601;
RA Xu D.-Q., Wu Y.-Q.;
RT "Cloning of the hup gene cluster from Rhodobacter sphaeroides 601.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=601;

```

RA Xu D.Q., Wu Y.Q.;
RT "Cloning, sequencing and function of the hupr gene from Rhodobacter
sphaeroides 601.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS A SIGMA-54 FACTOR INTERACTION ATP-BINDING
DOMAIN.

DR EMBL; AJ243734; CAC01253.1; -;
DR EMBL; AJ315650; CAC42238.1; -;
DR HSSP; P41789; INTR.

DR InterPro; IPR002197; HTH_Fis.

DR InterPro; IPR001289; Response_reg.

DR InterPro; IPR002078; Sig54_interact.

DR Pfam; PF02954; HTH_8; 1.

DR Pfam; PF00072; response_reg; 1.

DR Pfam; PF00158; Sigma54_activat; 1.

DR PRINTS; PR01590; HTHFIS.

DR ProDom; PD000039; Response_reg; 1.

DR SMART; SM00448; REC; 1.

DR TIGRFAMs; TIGR01199; HTH_Fis; 1.

DR PROSITE; PS00688; SIGMA54_INTERACT_3; UNKNOWN_1.

DR PROSITE; PS00645; SIGMA54_INTERACT_4; 1.

KW ATP-binding; DNA-binding; phosphorylation; sensory transduction;
transcription regulation.

KW SEQUENCE 491 AA; 54031 MW; 348616E053BBD5AE CRC64;

Query Match 0.7%; Score 8; DB 2; Length 491;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 LLLALLLV 420
|||||

DB 20 RMALEDEF 27

RESULT 59
Q9WUD0 PRELIMINARY; PRT; 491 AA.

ID Q9WUD0;
AC Q9WUD0; 01-NOV-1999 (TREMblrel. 12, Created)

DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)

DE Cytochrome P450 2B10 related protein.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=LIVER;

RA MEDLINE=99240492; PubMed=10222226;

RT "Isolation of a cyp2b10-like cDNA and of a clone derived from a
cyp2b10-like pseudogene.";

RL Biochem. Biophys. Res. Commun. 258:11-16(1999).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; AF128849; AAD28466.1; -;

DR HSSP; P00179; IDT6.

DR InterPro; IPR001128; Cytochrome_P450.

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.

KW Heme; Monooxygenase; Oxidoreductase.

KW SEQUENCE 491 AA; 55870 MW; A1338E108D77025E CRC64;

Query Match 0.7%; Score 8; DB 11; Length 491;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

DB

1134 LLLALLLV 1141
|||||

6 LLLALLLV 13

RESULT 60

Q64584 PRELIMINARY; PRT; 491 AA.

AC 064584;

DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)

DE Rat cytochrome P-450b (phenobarbital-inducible).

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=85234490; PubMed=2989270;

RA Suwa Y., Mizukami Y., Sogawa K., Fujii-Kuriyama Y.;

RT "Gene structure of a major form of phenobarbital-inducible cytochrome
P-450 in rat liver.";

RL J. Biol. Chem. 260:7980-7984(1985).

CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

DR EMBL; L00320; AAA41046.1; JOINED.

DR EMBL; L00313; AAA41046.1; JOINED.

DR EMBL; L00314; AAA41046.1; JOINED.

DR EMBL; L00315; AAA41046.1; JOINED.

DR EMBL; L00316; AAA41046.1; JOINED.

DR EMBL; L00317; AAA41046.1; JOINED.

DR EMBL; L00318; AAA41046.1; JOINED.

DR EMBL; L00319; AAA41046.1; JOINED.

DR HSSP; P00179; IDT6.

DR InterPro; IPR001128; Cytochrome_P450.

DR Pfam; PF00067; P450; 1.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.

KW Heme; Monooxygenase; Oxidoreductase.

KW SEQUENCE 491 AA; 56047 MW; E9144D72E522F0C CRC64;

Query Match 0.7%; Score 8; DB 11; Length 491;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
|||||

DB 6 LLLALLLV 13

RESULT 61

Q73804 PRELIMINARY; PRT; 497 AA.

AC 073804;

DT 01-AUG-1998 (TREMblrel. 07, Created)

DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)

DE Hypothetical 56.4 kDa protein.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

OC Tetraodontidae; Takifugu.

OX NCBI_TaxID=31033;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99177347; PubMed=10077531;

RA Gellner K., Brenner S.;

RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";

RL Genome Res. 9:251-258(1999).

DR EMBL; AF056116; AAC34385.1; -;

KW Hypothetical protein.

KW SEQUENCE 497 AA; 56376 MW; 802743E93EB93E64 CRC64;

Query Match 0.7%; Score 8; DB 13; Length 497;

Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLLLLV 1141
|||||||
Db 169 LLLLLLV 176

RESULT 62

Q9AS41 PRELIMINARY; PRT; 518 AA.

ID Q9AS41
AC Q9AS41:
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Putative cytochrome P450.
GN P0416G1.1 OR B111C09.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euphorbiaceae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0416G1.1";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT clone:B111C09.1";
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AP002968; BAB39252.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 518 AA; 57504 MW; 5A58E8246CC7ABC9 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 518;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLLLLV 1141
|||||||
Db 8 LLLLLLV 15

RESULT 63

Q8WVY3 PRELIMINARY; PRT; 573 AA.

ID Q8WVY3
AC Q8WVY3:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical 64.3 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=CERVIX;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC013741; AAH13741.1; -.
DR InterPro: IPR001909; KRAB.

DR InterPro: IPR000822; Znf_C2H2.
DR Pfam: PF00096; ZF-C2H2; 12.
DR PRINTS: PR00048; ZINC_FINGER.
DR ProDom: PD000003; Znf_C2H2; 10.
DR SMART: SM00355; Znf_C2H2; 12.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_12.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 12.
KW Hypothetical protein.
SQ SEQUENCE 573 AA; 64291 MW; 446498FAA8BE73E7 CRC64;

Query Match 0.7%; Score 8; DB 4; Length 573;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 898 FSCSSLS 905
|||||||
Db 270 FSCSSLS 277

RESULT 64

Q8R6N2 PRELIMINARY; PRT; 577 AA.

ID Q8R6N2
AC Q8R6N2:
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE ABC-type multidrug/protein/lipid transport system, ATPase
DE component. TTE2769.
GN MDL812 OR TTE2769.
OS Thermomicrobacter tengcongensis.
OC Bacteria; Firmicutes; Bacilli; Clostridium group; Clostridia;
OC Thermomicrobacteriales; Thermomicrobacteriaceae; Thermomicrobacter.
OX NCBI_TaxID=119072;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013214; AAM25873.1; -.
KW Complete proteome.
SQ SEQUENCE 577 AA; 66097 MW; BA6642DFF1BE7545 CRC64;

Query Match 0.7%; Score 8; DB 16; Length 577;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 EEVVRRAK 233
|||||||
Db 429 EEVVRRAK 436

RESULT 65

Q8TRM8 PRELIMINARY; PRT; 596 AA.

ID Q8TRM8
AC Q8TRM8:
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE F420-nonreducing hydrogenase.
GN VNR1 OR MAl147.
OS Methanobrevibacterium acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=1193238;

RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atchord D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., DeRellano K., Johnson R.,
 RA Linton L., McKernan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leight J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umagay L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.,
 RT "The genome of *Mechanosarcina acetivorans* reveals extensive metabolic
 RT and physiological diversity." ;
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE010781; AAM04568.1; -.
 KW Complete proteome.
 SQ SEQUENCE 596 AA; 65478 MW; 3780AFC2448F8D5B CRC64;

Query Match 0.7%; Score 8; DB 17; Length 596;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 SVLQHWG 51
 DB 187 SVLQHWG 194

RESULT 66
 ID Q39775 PRELIMINARY; PRT; 607 AA.
 AC Q39775;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Legumin, 11S globulin.
 OS Gnetum gnetum.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Gnecotphyta; Gnecopsida; Gnecatales; Gnecaceae; Gnetum.
 OX NCBI_Taxid=3382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SEED;
 RX MEDLINE=98440765; PubMed=9767693;
 RA Shucov A.D., Braun H., Chesnokov Y.V., Horstmann C., Kakhovskaya I.A.,
 RA Baumlein H.;
 RT "Sequence peculiarity of Gnetalean legumin-like seed storage
 RT proteins." ;
 RL J. Mol. Evol. 47:486-492(1998).
 DR EMBL: Z50779; CAA90642.1; -.
 DR InterPro: IPR000459; Seedstore 11S.
 DR Pfam: PF00190; Seedstore 11S; 1.
 DR PRINTS: PR00439; 11SGLOBULIN.
 SQ SEQUENCE 607 AA; 68910 MW; 7AD6DDC604FB3A60 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 607;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 10 LLLALLV 17

RESULT 67
 ID Q31615 PRELIMINARY; PRT; 614 AA.
 AC Q31615;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE YUBQ protein.
 OS YUBQ.
 GN Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Bourlier L., Brans A., Braun M., Biggelli S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Deutz F., Devine K.M., Distenfeld A., Ehrlich S.D., Emerson P.T.,
 RA Euteneier K.D., Eyring J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaetr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Puig P., Purnelle D., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Mamuit R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium *Bacillus*
 RT *subtilis*." ;
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: Z99110; CAB13021.1; -.
 DR InterPro: IPR000676; Nah_Ekxngtr.
 DR InterPro: IPR000309; TrkA_Kupcake.
 DR InterPro: IPR003148; TrkA_N.
 DR Pfam: PF00939; Na_H_Exchange; 1.
 DR Pfam: PF02080; TrkA-C; 1.
 DR Pfam: PF02254; TrkA-N; 1.
 KW Complete proteome.
 SQ SEQUENCE 614 AA; 67467 MW; EAAAE1850799163B CRC64;

Query Match 0.7%; Score 8; DB 16; Length 614;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 313 LLLALLV 320

RESULT 68
 ID Q91570 PRELIMINARY; PRT; 622 AA.
 AC Q91570;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Integrin alpha5tr subunit (Fragment).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

OC Xenopodinae; Xenopus.
OK NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=95344994; PubMed=7619730;
RA Joos T.O., Whitlaker C.A., Meng F., Desimone D.W., Gnan V., Hansen P.,
RT "Integrin alpha 5 during early development of Xenopus laevis.";
RL Mech. Dev. 50:187-199(1995).
DR EMBL; U12646; AAA9667.1; -.
DR HSSP; P06756; IJY2.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP; 1.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int_alpha; 1.
FT NON TER 1 1
SQ SEQUENCE 622 AA; 69003 MW; 78090D59A6B3660C CRC64;

Query Match 0.7%; Score 8; DB 13; Length 622;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
Db 604 GLLALL 611

RESULT 69
017795 PRELIMINARY; PRT; 634 AA.
AC 017795;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hypothetical 68.2 kDa protein.
GN C07G1.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hawkins J.;
RT "The sequence of C. elegans cosmid C07G1.";
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U58751; AAB00657.2; -
DR InterPro; IPR000095; PARbox/Rhobndg.
DR InterPro; IPR000697; Ranbp1_MASP.
DR InterPro; IPR001960; WH1.
DR InterPro; IPR003124; WH2.
DR Pfam; PF00786; PBD; 2.
DR Pfam; PF00568; WH1; 2.
DR Pfam; PF02205; WH2; 4.
DR PRINTS; PR01582; KV3CHANNEL.
DR SMART; SM00285; PBD; 2.
DR SMART; SM00461; WH1; 2.
DR SMART; SM00246; WH2; 2.
DR PROSITE; PS50108; GBD; 2.

KW Hypothetical protein.
SQ SEQUENCE 634 AA; 68201 MW; C9E489254C070DD0 CRC64;

Query Match 0.7%; Score 8; DB 5; Length 634;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 672 BAVCLTAA 679
Db 48 BAVCLTAA 55

RESULT 70
09CWM6 PRELIMINARY; PRT; 636 AA.
ID 09CWM6
AC 09CWM6;
DT 01-JUN-2001 (TRENBLREL. 17, Created)
DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Hypothetical protein PM0698.
GN PM0698.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.U., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.,
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006106; AAK02782.1; -.
DR InterPro; IPR001091; CNA_Mettransf.
DR InterPro; IPR002295; D21N6_mltfrase.
DR InterPro; IPR002941; N6/N4_Mtase.
DR InterPro; IPR002052; N6_Mtase.
DR Pfam; PF01555; N6_N4_Mtase; 1.
DR PRINTS; PR00506; D21N6MTFRASE.
DR PRINTS; PR00508; S21N4MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 636 AA; 72597 MW; D3A20817BCE8A6A CRC64;

Query Match 0.7%; Score 8; DB 16; Length 636;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 969 KTLRLVQN 976
Db 343 KTLRLVQN 350

RESULT 71
070491 PRELIMINARY; PRT; 670 AA.
ID 070491
AC 070491;
DT 01-AUG-1998 (TRENBLREL. 07, Created)
DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)
DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
DE Retinoic acid-responsive protein.
GN STRA6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95377533; PubMed=7649373;
RA Boulliet P., Oulad-Abdelghani M., Vicaire S., Garnier J.M.,
RT "Efficient cloning of cDNAs of retinoic acid-responsive genes in P19
embryonal carcinoma cells and characterization of a novel mouse gene,

RT Stral (mouse LERK-2/Ed192).";
 RL Dev. Biol. 170:420-433 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97346723; PubMed=92031140;
 RA Boulliet P., Sapin V., Chazaud C., Messadeg N., Decimo D., Dolle P.,
 RA Chambon P.;
 RA "Developmental expression pattern of Stral6, a retinoic acid-responsive
 RT gene encoding a new type of membrane protein.";
 RL Mech. Dev. 63:173-186 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Boulliet P.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF062476; AAC16016.1; -
 DR MGD; MGI:107742; Stral6
 SQ SEQUENCE 670 AA; 73774 MW; 6AC669F3EE16A4F7 CRC64;

Query Match 0.7%; Score 8; DB 11; Length 670;
 Best Local Similarity 100.0%; Pred. No. 94;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 64 LLLALLV 71

RESULT 72

ID 082985 PRELIMINARY; PRT; 685 AA.
 AC 082985;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Ferrichrome transport protein PhnB.
 OS Sty0221.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxId=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*
 RT enterica serovar typhi CT18.";
 RL Nature 413:848-852 (2001).
 DR EMBL; AL627265; CAD01355.1; -
 DR InterPro; IPR000522; Feccd.
 DR Pfam; PF01032; Feccd; 2.
 DR ProDom; PD001557; Feccd; 1.
 KW Complete proteome.
 SQ SEQUENCE 685 AA; 73148 MW; F72A8ABE03556B4D CRC64;

Query Match 0.7%; Score 8; DB 16; Length 685;
 Best Local Similarity 100.0%; Pred. No. 96;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
 DB 36 LLLALLV 43

RESULT 73

Q96FV1

ID Q96FV1 PRELIMINARY; PRT; 704 AA.
 AC Q96FV1;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE KIAA1937 protein (Fragment).
 GN KIAA1937.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRN;
 RX MEDLINE=21456161; PubMed=11572484;
 RA Nagase T., Kiruno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XXI.
 RT large proteins.";
 RT DNA Res. 8:179-187 (2001).
 DR EMBL; AB067524; BAB67830.1; -
 FT NON_TER 1 1
 SQ SEQUENCE 704 AA; 82013 MW; AFG8E675A2CC8BEC CRC64;

Query Match 0.7%; Score 8; DB 4; Length 704;
 Best Local Similarity 100.0%; Pred. No. 98;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1130 VIIGLLLL 1137
 DB 1 VIIGLLLL 8

RESULT 74

ID Q91N19 PRELIMINARY; PRT; 739 AA.
 AC Q91N19;
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE VP3.
 OS Kadipiro virus.
 OC Viruses; dsRNA viruses; Reoviridae; Coltivirus.
 OX NCBI_TaxId=104580;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KT-7075;
 RX MEDLINE=20273996; PubMed=10811934;
 RA Attoui H., Billoir F., Biagini P., de Micco P., de Lamballerie X.;
 RT "Complete sequence determination and genetic analysis of Banna virus
 RT and Kadipiro virus: proposal for assignment to a new genus
 RT (Seadornavirus) within the family Reoviridae.";
 RL J. Gen. Virol. 81:1507-1515 (2000).
 DR EMBL; AF134510; AAF78851.1; -
 SQ SEQUENCE 739 AA; 85527 MW; A651D16C33A5709 CRC64;

Query Match 0.7%; Score 8; DB 12; Length 739;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1125 ILIGSVLG 1132
 DB 638 ILIGSVLG 645

RESULT 75

ID Q9CAD3 PRELIMINARY; PRT; 767 AA.
 AC Q9CAD3;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Putative UDP-N-acetylmuramoylalanine-D-glutamate--2, 6-diaminopimelate

DE ligase.
GN F24D7.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OK NCBI_taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=1130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.U., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luvos J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nietman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana." Nature 408:816-820(2000).
RL Nature 408:816-820(2000).
DR EMBL; AC011622; AAG52413.1; -.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR Pfam; PF01225; Mur_ligase; 1.
DR Pfam; PF02875; Mur_ligase_C; 1.
SQ SEQUENCE 767 AA; 85022 MW; 93887DF10F50C748 CRC64;

Query Match 0.7%; Score 8; DB 10; Length 767;
Best Local Similarity 100.0%; Pred.No.1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 AVGAPLED 575
|||||||
Db 559 AVGAPLED 566

Search completed: July 16, 2003, 08:33:40
Job time : 58 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:25:34 ; Search time 46 Seconds
(without alignments)
3380.510 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167

Sequence: 1 MELPVTHTLPLVFLTGLC.....GFPAHKKIPBEKEKLEQ 1167

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size: 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database:

A.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1167	100.0	1167	21	AAV32242 Human integrin sub
2	1152	98.7	1152	22	AAAB6457 Human secreted pro
3	973	83.4	1132	21	AAV32243 Human integrin sub
4	427	36.6	1152	22	AAAB6458 Human secreted pro
5	427	36.6	1167	22	AAAB6458 Human secreted pro
6	195	16.7	195	23	AAU76853 Human integrin alp
7	192	16.5	195	23	AAU76862 Human integrin alp
8	122	10.5	124	20	AAV41735 Human PRO827 prote
9	122	10.5	124	21	AAAB4291 Human PRO827 (UNQ4
10	122	10.5	124	22	AAU29100 Human PRO polypept

11	122	10.5	124	23	ABB95485 Human angiogenesis
12	122	10.5	124	23	ABB84879 Human PRO827 prote
13	42	3.6	303	21	AAV32282 Mouse integrin sub
14	22	1.9	22	21	AAV32244 Human integrin sub
15	14	1.2	18	21	AAV32245 Bovine integrin su
16	13	1.1	103	22	AAU76757 Human novel centr
17	13	1.1	103	22	AAU19822 Human novel extrac
18	13	1.1	103	23	ABP48042 Human novel extrac
19	13	1.1	148	22	AAU19634 Human novel extrac
20	13	1.1	148	22	AAU19794 Human novel extrac
21	13	1.1	148	23	ABP47854 Human polypeptide
22	13	1.1	148	23	ABP48014 Human polypeptide
23	13	1.1	193	23	AAU76854 Human integrin alp
24	13	1.1	193	23	AAU76854 Human integrin alp
25	13	1.1	193	23	AAU76851 Human integrin alp
26	13	1.1	195	23	AAU76860 Human integrin alp
27	13	1.1	214	22	AAU76860 Human integrin alp
28	13	1.1	214	22	AAU76860 Human integrin alp
29	13	1.1	707	22	AAU19663 Human alpha1 integ
30	13	1.1	707	22	AAU19663 Human novel extrac
31	13	1.1	1034	21	AAU76863 Human polypeptide
32	13	1.1	1179	23	AAU76863 Human polypeptide
33	13	1.1	1180	23	ABP47883 Human novel extrac
34	13	1.1	1188	22	AAU14231 Human novel protei
35	13	1.1	1188	22	AAU14467 Human novel protei
36	13	1.1	1188	22	AAU14467 Human novel protei
37	13	1.1	1188	22	AAU14467 Human novel protei
38	13	1.1	1188	22	AAU14467 Human novel protei
39	13	1.1	1188	22	AAU14467 Human novel protei
40	13	1.1	1188	22	AAU14467 Human novel protei
41	13	1.1	1189	21	AAU14467 Human novel protei
42	13	1.1	1189	22	AAU14467 Human novel protei
43	12	1.0	1183	20	AAU14467 Human novel protei
44	11	0.9	117	22	ABG12950 Human novel protei
45	11	0.9	185	22	AAU09125 Human novel protei
46	11	0.9	545	23	ABP72288 Human novel protei
47	11	0.9	688	23	ABP72300 Human novel protei
48	11	0.9	696	23	ABP72288 Human novel protei
49	11	0.9	697	19	AAW98238 Human novel protei
50	11	0.9	823	22	ABG25584 Human novel protei
51	11	0.9	979	22	ABG29239 Human novel protei
52	11	0.9	1183	20	AAU07729 Human novel protei
53	11	0.9	1367	19	AAW70542 Human novel protei
54	10	0.9	1178	16	AAW82656 Human novel protei
55	10	0.9	1223	22	ABG05837 Human novel protei
56	9	0.8	123	21	AAU32249 Human novel protei
57	9	0.8	1065	18	AAW35852 Human novel protei
58	9	0.8	1086	20	AAW81839 Human novel protei
59	9	0.8	1145	21	AAW81839 Human novel protei
60	9	0.8	1170	11	AAW81839 Human novel protei
61	9	0.8	1170	16	AAW81839 Human novel protei
62	9	0.8	1170	16	AAW81839 Human novel protei
63	9	0.8	1170	23	ABP76377 Human novel protei
64	9	0.8	1170	23	AAU80251 Human novel protei
65	8	0.7	8	21	AAV32248 Human novel protei
66	8	0.7	17	20	AAW99005 Human novel protei
67	8	0.7	17	20	AAV12862 Human novel protei
68	8	0.7	66	21	AAW00326 Human novel protei
69	8	0.7	67	20	AAV41696 Human novel protei
70	8	0.7	67	21	AAW44252 Human novel protei
71	8	0.7	67	21	AAW44252 Human novel protei
72	8	0.7	67	21	AAW44252 Human novel protei
73	8	0.7	67	21	AAW44252 Human novel protei
74	8	0.7	67	21	AAW44252 Human novel protei
75	8	0.7	67	21	AAW44252 Human novel protei
76	8	0.7	67	21	AAW44252 Human novel protei
77	8	0.7	67	21	AAW44252 Human novel protei
78	8	0.7	67	21	AAW44252 Human novel protei
79	8	0.7	67	21	AAW44252 Human novel protei
80	8	0.7	67	21	AAW44252 Human novel protei
81	8	0.7	67	21	AAW44252 Human novel protei
82	8	0.7	67	21	AAW44252 Human novel protei
83	8	0.7	67	21	AAW44252 Human novel protei

84	8	0.7	104	18	AAW07680	Human LST-1 (leuko
85	8	0.7	167	22	AAE04539	Human cytokine, ZC
86	8	0.7	179	21	AAE04539	Human IL-19/AE289
87	8	0.7	179	21	AAE04539	Human T cell induc
88	8	0.7	179	22	AAE04539	Human T cell deriv
89	8	0.7	179	22	AAE04539	Human cytokine, ZC
90	8	0.7	179	22	AAE04539	Human IL-TIF polyp
91	8	0.7	179	22	AAE04539	Human IL-TIF polyp
92	8	0.7	179	22	AAE04539	Human IL-TIF polyp
93	8	0.7	179	22	AAE04539	Human IL-TIF polyp
94	8	0.7	179	22	AAE04539	Human IL-TIF polyp
95	8	0.7	179	22	AAE04539	Human IL-TIF polyp
96	8	0.7	179	22	AAE04539	Human IL-TIF polyp
97	8	0.7	179	22	AAE04539	Human IL-TIF polyp
98	8	0.7	179	22	AAE04539	Human IL-TIF polyp
99	8	0.7	179	22	AAE04539	Human IL-TIF polyp
100	8	0.7	179	22	AAE04539	Human IL-TIF polyp

ALIGNMENTS

RESULT 1

AA32242
ID AAY32242 standard; Protein; 1167 AA.

AC	XX	AA32242;				
XX	XX	15-FEB-2000	(first entry)			
DT	XX	Human integrin subunit alpha-10.				
XX	XX	Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;				
XX	XX	osteoarthritis; osteoarthritis; cancer; atherosclerosis;				
XX	XX	inflammation; therapy; cartilage; chondrocyte; osteoblast;				
XX	XX	fibroblast; vaccine; marker.				
XX	XX	Homo sapiens.				
OS	XX	Location/Qualifiers				
XX	XX	1..22				
XX	XX	/note="signal peptide"				
XX	XX	23..1145				
XX	XX	/note="mature protein"				
XX	XX	23..1120				
XX	XX	/note="extracellular domain"				
XX	XX	1121..1145				
XX	XX	/note="transmembrane domain"				
XX	XX	1122..1167				
XX	XX	/note="cytoplasmic domain, specifically claimed				
XX	XX	in Claim 21"				
XX	XX	162..359				
XX	XX	/note="I-domain"				
XX	XX	494..502				
XX	XX	/note="cation binding site motif"				
XX	XX	558..566				
XX	XX	/note="cation binding site motif"				
XX	XX	620..628				
XX	XX	/note="cation binding site motif"				
XX	XX	98				
XX	XX	/note="N-glycosylated"				
XX	XX	336				
XX	XX	/note="N-glycosylated"				
XX	XX	364				
XX	XX	/note="N-glycosylated"				
XX	XX	733				
XX	XX	/note="N-glycosylated"				
XX	XX	839				
XX	XX	/note="N-glycosylated"				
XX	XX	921				
XX	XX	/note="N-glycosylated"				
XX	XX	1018				
XX	XX	/note="N-glycosylated"				

FT	Modified-site	1039	/note="N-glycosylated"
FT	FT		
XX	XX	WO9951639-A1.	
XX	XX	14-OCT-1999.	
XX	XX	31-MAR-1999;	99WO-SE00544.
XX	XX	02-APR-1998;	98SE-0001164.
XX	XX	28-JAN-1999;	99SE-0000319.
XX	XX	(ACTI-) ACTIVE BIOTECH AB.	
XX	XX	Lundgren-Akerlund E;	
XX	XX	WPI: 2000-052639/04.	
XX	XX	N-PDB; AA234719.	
XX	XX	New isolated integrin subunit alpha-10, used as a marker or target	
XX	XX	molecule for cells during development, regeneration and pathological	
XX	XX	conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or	
XX	XX	inflammation	
XX	XX	Claim 1; Fig 6; 90pp; English.	
XX	XX	This sequence represents novel human chondrocyte integrin subunit	
XX	XX	alpha-10 (ISa10). A splice variant is given in AAY32243. The	
XX	XX	invention relates to a recombinant or isolated integrin heterodimer	
XX	XX	comprising the alpha10 subunit in association with subunit beta	
XX	XX	(especially beta-1). The heterodimer and the subunit alpha-10 can	
XX	XX	be used as markers or targets of all types of cells, e.g. of	
XX	XX	chondrocytes, osteoblasts and fibroblasts. They can also be used:	
XX	XX	for treating pathological conditions involving ISa10, such as	
XX	XX	damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;	
XX	XX	for detecting the formation of cartilage during embryonal	
XX	XX	development, physiological or therapeutic repair of cartilage,	
XX	XX	or detecting regeneration of cartilage or chondrocytes during	
XX	XX	transplantation of cartilage or chondrocytes; for selection and	
XX	XX	analysis or for sorting, isolating or purification of chondrocytes	
XX	XX	and for in vitro studies of differentiation of chondrocytes; and as	
XX	XX	a target for anti-adhesive drugs or molecules in tendon, ligament,	
XX	XX	skeletal muscle or other tissues where adhesion impairs the function	
XX	XX	of the tissue (all claimed). ISa10 binding entities can be used to	
XX	XX	determine the differentiation-state of cells during embryonic	
XX	XX	development, angiogenesis or development of cancer, in pathological	
XX	XX	conditions such as rheumatoid arthritis, osteoarthritis or cancer,	
XX	XX	in tissue regeneration or in therapeutic and physiological repair	
XX	XX	of cartilage (claimed). A vaccine comprising the integrin	
XX	XX	heterodimer or subunit alpha-10 is also claimed. ISa10	
XX	XX	polynucleotides, vectors, host cells and methods of producing	
XX	XX	recombinant ISa10 are also claimed.	
XX	XX	Sequence 1167 AA;	
XX	XX	Query Match 100.0%; Score 1167; DB 21; Length 1167;	
XX	XX	Best Local Similarity 100.0%; Pred. No. 0;	
XX	XX	Matches 1167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
XX	XX	1 MELPVTHLFLPLVFLTGLCPNLDENHPLFPQPPAEFGYSLVQHVGGGRWMLVGA 60	
XX	XX	1 MELPVTHLFLPLVFLTGLCPNLDENHPLFPQPPAEFGYSLVQHVGGGRWMLVGA 60	
XX	XX	61 FMDGSGDRGDVYRCFPGAHNPPCAKGHIGDVLQNGSSHPAIVMHLGMSLLETDDGG 120	
XX	XX	61 FMDGSGDRGDVYRCFPGAHNPPCAKGHIGDVLQNGSSHPAIVMHLGMSLLETDDGG 120	
XX	XX	61 FMDGSGDRGDVYRCFPGAHNPPCAKGHIGDVLQNGSSHPAIVMHLGMSLLETDDGG 120	
XX	XX	121 FMACAPLWSRACGSSVFSSGICARVDASFOQGSIAPTAORCPTMYDVIVLQDSNSIYP 180	
XX	XX	121 FMACAPLWSRACGSSVFSSGICARVDASFOQGSIAPTAORCPTMYDVIVLQDSNSIYP 180	
XX	XX	121 FMACAPLWSRACGSSVFSSGICARVDASFOQGSIAPTAORCPTMYDVIVLQDSNSIYP 180	
XX	XX	181 MSEVOTFARRLVGKLFIDPEIQVGLVQYGSPPHENSIGFRTKEEVYRAAKLSREG 240	

```

Db      161 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEMSIGDFRTKEEVVRAAKNLSRREG 240
QY      241 RETKTAQAIMVACTEGFSQSHGSRPEARLLVVTVDGSHDEELPALAKACEAGRVTY 300
Db      241 RETKTAQAIMVACTEGFSQSHGSRPEARLLVVTVDGSHDEELPALAKACEAGRVTY 300
QY      301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNNVTDEALTDIIVALGDRIFGLEGS 360
Db      301 GIAVLGHLRRQRDPSSFLREIRTIASDPDERFFNNVTDEALTDIIVALGDRIFGLEGS 360
QY      361 HAENSSFGLEMSQIGFSTHRLKDGILRGMGAYVMGSGVLMLEGGHLLPFRMALLEDEF 420
Db      361 HAENSSFGLEMSQIGFSTHRLKDGILRGMGAYVMGSGVLMLEGGHLLPFRMALLEDEF 420
QY      421 PPALONHAAYLGSYSSMLRGRLFLSGAPFRHGRVIAFOLKKGDAVVAOSLQGE 480
Db      421 PPALONHAAYLGSYSSMLRGRLFLSGAPFRHGRVIAFOLKKGDAVVAOSLQGE 480
QY      481 QIGSYFGSELCPDIDRDGTTDVLVVAAPMFLGPONKETRIVYVYLVGQOSILLTQGTLO 540
Db      481 QIGSYFGSELCPDIDRDGTTDVLVVAAPMFLGPONKETRIVYVYLVGQOSILLTQGTLO 540
QY      541 PEPPODAFFGFAMGALPDLDNODGFADVAVGALBEGHOGALYLYHGTSQVAPRPHAOGLA 600
Db      541 PEPPODAFFGFAMGALPDLDNODGFADVAVGALBEGHOGALYLYHGTSQVAPRPHAOGLA 600
QY      601 AASMFHALSYFGSRVDGRDLDDDLVDVAVAGAGAAIILSSRPVHLTPSLLEVTPOALS 660
Db      601 AASMFHALSYFGSRVDGRDLDDDLVDVAVAGAGAAIILSSRPVHLTPSLLEVTPOALS 660
QY      661 VVQDRCRRGQAVCLTALCFQVTSRTPGRMDFYMRFTASLDEMTAGARAAFDGSGQ 720
Db      661 VVQDRCRRGQAVCLTALCFQVTSRTPGRMDFYMRFTASLDEMTAGARAAFDGSGQ 720
QY      721 RLSPRLKLSVANNVCEQHFHVLDTSYLRVALTVTFALDNTTKPGPVLENGSPSTIQ 780
Db      721 RLSPRLKLSVANNVCEQHFHVLDTSYLRVALTVTFALDNTTKPGPVLENGSPSTIQ 780
QY      781 KLVPSSKDCGPNCEVTDLVQVNMDIRSKAPFVNGRGRKLVSTLENRKENAYNT 840
Db      781 KLVPSSKDCGPNCEVTDLVQVNMDIRSKAPFVNGRGRKLVSTLENRKENAYNT 840
QY      841 SLSTIFSRNLHLASLTPORESPIKYECAPSAHARLCSVGHVPFOTGAKVTFLLFEFSC 900
Db      841 SLSTIFSRNLHLASLTPORESPIKYECAPSAHARLCSVGHVPFOTGAKVTFLLFEFSC 900
QY      901 SSLSQVQKGLTASSDSLENRGTLOENTRQTSAYIQYEPHLLFSSSESTLHREYVHPYGL 960
Db      901 SSLSQVQKGLTASSDSLENRGTLOENTRQTSAYIQYEPHLLFSSSESTLHREYVHPYGL 960
QY      961 PVGPGEFFKTLRVONLGGCYVSGLIISALLPAVHAGGNYFLLSOVITNNASCIYONLT 1020
Db      961 PVGPGEFFKTLRVONLGGCYVSGLIISALLPAVHAGGNYFLLSOVITNNASCIYONLT 1020
QY      1021 EPPGPVHPEELQHTNRNGSNTOQOVVRCHLGOLAKTEVSVGLRLVHNEFFRRAFK 1080
Db      1021 EPPGPVHPEELQHTNRNGSNTOQOVVRCHLGOLAKTEVSVGLRLVHNEFFRRAFK 1080
QY      1081 SITVSTELTEGTEBSSVLOLTASRWSSELEEVOTRPLISLMLLIGSVLGGILLALL 1140
Db      1081 SITVSTELTEGTEBSSVLOLTASRWSSELEEVOTRPLISLMLLIGSVLGGILLALL 1140
QY      1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167
Db      1141 VFCLMKLGFPAHKKIPEEKREKLEQ 1167

```

```

RESULT 2
AAB64657
ID AAB64657 standard; Protein; 1152 AA.
XX
AC AAB64657;
XX

```

```

DT      22-MAR-2001 (first entry)
XX
DE      Human secreted protein BLAST search protein SEQ ID NO: 167.
XX
KW      Cyclostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW      antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW      vulnerability; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW      cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW      neurological disease; infection; human; secreted protein.
XX
OS      Homo sapiens.
XX
PN      WO200077197-A1.
XX
PD      21-DEC-2000.
XX
PF      01-JUN-2000; 2000WO-US14934.
XX
PR      11-JUN-1999; 99US-0138599.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX      (ROSE/) ROSEN C A.
XX
PI      Rosen CA, Ruben SM, Komatsoulis GA.
XX
DR      WPI; 2001-032312/04.
XX
PT      Isolated nucleic acid molecule encoding a human secreted protein is
XX      used in preventing, treating or ameliorating a medical condition -
XX
PS      Disclosure; Page 543-546; 558pp; English.
XX
CC      The invention relates to the isolation of genes AAF32757-F32803 encoding
CC      the human secreted proteins AAB64549-B64594. The sequence is a search
CC      result from a BLASTX homology search. The genes and proteins are useful
CC      for preventing, ameliorating or treating medical conditions, e.g. by
CC      protein or gene therapy. The genes are isolated from a range of human
CC      tissues disclosed in the specification. The nucleic acids, proteins,
CC      antibodies and (ant)agonists are useful in the diagnosis, treatment
CC      and prevention of: (a) cancer; e.g. breast and ovarian cancer, and
CC      other cancers of the adrenal gland, bone, bone marrow, breast,
CC      gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC      disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC      anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC      multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC      cardiovascular disorders such as myocardial ischaemias; (d) wound
CC      healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC      and (f) infectious diseases such as viral, bacterial, fungal and
CC      parasitic infections.
XX
SQ      Sequence 1152 AA;
XX
Query Match 98.7%; Score 1152; DB 22; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 MELPFVTHLFLPLVFLTGLSGPFLNDEHHPFLPFGPPAEHGYVLOHVGSGGORMLVGA 60
Db      1 MELPFVTHLFLPLVFLTGLSGPFLNDEHHPFLPFGPPAEHGYVLOHVGSGGORMLVGA 60
QY      61 PWDPSGDRRDYVRCVPGAGAHNAPCAKGLHGDYOLQNGSSHPANMHLGMSLLETGDDGG 120
Db      61 PWDPSGDRRDYVRCVPGAGAHNAPCAKGLHGDYOLQNGSSHPANMHLGMSLLETGDDGG 120
QY      121 FMACAPLMSRACGSSVPSGICARVDASFOPOGSLAPTAORCPYMDVYVYLDGNSNITYP 180
Db      121 FMACAPLMSRACGSSVPSGICARVDASFOPOGSLAPTAORCPYMDVYVYLDGNSNITYP 180
QY      181 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEMSIGDFRTKEEVVRAAKNLSRREG 240
Db      181 MSEVOTFLRLVGLFIDPEIQIUGLVQYGSPPVHEMSIGDFRTKEEVVRAAKNLSRREG 240
QY      241 RETKTAQAIMVACTEGFSQSHGSRPEARLLVVTVDGSHDEELPALAKACEAGRVTY 300

```

```

DB      241 RETKTAQAIWVACTEBSQSHGCRPEAARLLVVTGSESDGELPAALKACAGAVTRY      300
QY      301 GTAVLGHVRRORDPSSFLREIRTTASDPERFEFNVTEBAATDIDVDLGDRIEGLSES      360
DB      301 GTAVLGHVRRORDPSSFLREIRTTASDPERFEFNVTEBAATDIDVDLGDRIEGLSES      360
QY      361 HANESSFGLSEMSQIGFSTRRLKDGILFGWVGAVDWGSVLMEGSHLPFPPMALEDEF      420
DB      361 HANESSFGLSEMSQIGFSTRRLKDGILFGWVGAVDWGSVLMEGSHLPFPPMALEDEF      420
QY      421 PPALQNHAAVLYGYSVSSMLRGGRULFLSGAPFRHRKVIYAFQKKDQAVRVAQSLQSE      480
DB      421 PPALQNHAAVLYGYSVSSMLRGGRULFLSGAPFRHRKVIYAFQKKDQAVRVAQSLQSE      480
QY      481 QIGSYFSGSELCPLDPTDRDGTDDVLAAPMFLGPONKERTGRVYVYVGOQSLTLTGCTIQ      540
DB      481 QIGSYFSGSELCPLDPTDRDGTDDVLAAPMFLGPONKERTGRVYVYVGOQSLTLTGCTIQ      540
QY      541 PEPPDARFGFAMGALPDINODGFADAVAGAPLEDHOGALYVHGTOGVRPHPAQRIA      600
DB      541 PEPPDARFGFAMGALPDINODGFADAVAGAPLEDHOGALYVHGTOGVRPHPAQRIA      600
QY      601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGOGAAILLSRPVHLTPSLEVPQAIS      660
DB      601 AASMPHALSYFGRSVDGRDLDDGDDLVDAVAGOGAAILLSRPVHLTPSLEVPQAIS      660
QY      661 VVQDRRRGOEAVCTALACFOVTSRTGMRDHQYMFRTASLDEMTGARAAPDQSGQ      720
DB      661 VVQDRRRGOEAVCTALACFOVTSRTGMRDHQYMFRTASLDEMTGARAAPDQSGQ      720
QY      721 RLSPRRRLSVGNVCEQLHFVLDTSYLRPVALTPTPALTDTTTPKGVNLEGSPTSIO      780
DB      721 RLSPRRRLSVGNVCEQLHFVLDTSYLRPVALTPTPALTDTTTPKGVNLEGSPTSIO      780
QY      781 KLVFESKDCGPDNECTDLVLQNMDIRSRKAPFVVRGGRVUVSTLLENKEMAYNT      840
DB      781 KLVFESKDCGPDNECTDLVLQNMDIRSRKAPFVVRGGRVUVSTLLENKEMAYNT      840
QY      841 SLSIISRNHLASLTPORESPIKVCAPSAHARLCSVGHVPFGATKATFLPEFSC      900
DB      841 SLSIISRNHLASLTPORESPIKVCAPSAHARLCSVGHVPFGATKATFLPEFSC      900
QY      901 SLSLQGVFCKLTASSDSLEENGTLOENTAOQTSAYIYEPHLLFSSSESTLHRYEVPYGTL      960
DB      901 SLSLQGVFCKLTASSDSLEENGTLOENTAOQTSAYIYEPHLLFSSSESTLHRYEVPYGTL      960
QY      961 PVGPGPEFTTLRVQNLGCVYVSGLIISALLPAVHAGNVFSLSGVITNNASCIVQNT      1020
DB      961 PVGPGPEFTTLRVQNLGCVYVSGLIISALLPAVHAGNVFSLSGVITNNASCIVQNT      1020
QY      1021 EPPGPVHPEELQHTRLNGSNTOCCOVACHLQOLAKGTEVSGLLRLVHNEFFPRRAXK      1080
DB      1021 EPPGPVHPEELQHTRLNGSNTOCCOVACHLQOLAKGTEVSGLLRLVHNEFFPRRAXK      1080
QY      1081 SLTVVSTFELGTEGSGVLQTEASRNSESLLEVQTRPILISLWILIGSVLGLLLALL      1140
DB      1081 SLTVVSTFELGTEGSGVLQTEASRNSESLLEVQTRPILISLWILIGSVLGLLLALL      1140
QY      1141 VFCIMKLGFFAH      1152
DB      1141 VFCIMKLGFFAH      1152

```

RESULT 3

AAV32243
ID AAV32243 standard; Protein; 1132 AA.

AC AAV32243;

DT 15-FEB-2000 (first entry)

XX Human integrin subunit alpha-10 splice variant.

```

XX      XX      Integrin alpha-10; 1Sa10; human; trauma; rheumatoid arthritis;
KW      KW      osteoarthritis; osteoarthritis; cancer; atherosclerosis;
KW      KW      inflammation; therapy; cartilage; chondrocyte; osteoblast;
KW      KW      fibroblast; vaccine; marker; splice variant.
OS      OS      Homo sapiens.
FH      FH      Key
FT      FT      Peptide
FT      FT      Protein
FT      FT      /note= "signal peptide"
FT      FT      /note= "mature protein"
PN      PN      WO951639-A1.
PD      PD      14-OCT-1999.
PF      PF      31-MAR-1999; 99WO-SE00544.
PR      PR      02-APR-1998; 98SE-0001164.
PR      PR      28-JAN-1999; 99SE-0000319.
XX      XX      (ACT1-) ACTIVE BIOTECH AB.
PA      PA      Lundgren-Akerlund E;
PI      PI      WPI: 2000-052639/04.
DR      DR      N-PSDB; AA234720.
XX      XX      New isolated integrin subunit alpha-10, used as a marker or target
PT      PT      molecule for cells during development, regeneration and pathological
PT      PT      conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT      PT      inflammation -
XX      XX      Claim 1; Page 43-48; 90pp; English.
PS      PS      This sequence represents a splice variant of novel human
CC      CC      chondrocyte integrin subunit alpha-10 (1Sa10). It is identical to
CC      CC      1Sa10 (see AAV32242) except for deletion of amino acids 975-986. The
CC      CC      invention relates to a recombinant or isolated integrin heterodimer
CC      CC      comprising the alpha10 subunit in association with subunit beta
CC      CC      (especially beta-1). The heterodimer, subunit alpha-10 or splice
CC      CC      variant can be used as a marker or target of all types of cells, e.g.
CC      CC      of chondrocytes, osteoblasts and fibroblasts. They can also be used:
CC      CC      for treating pathological conditions involving 1Sa10, such as
CC      CC      damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis;
CC      CC      for detecting the formation of cartilage during embryonal
CC      CC      development, physiological or therapeutic repairation of cartilage,
CC      CC      or detecting regeneration of cartilage or chondrocytes during
CC      CC      transplantation of cartilage or chondrocytes; for selection and
CC      CC      analysis or for sorting, isolating or purification of chondrocytes;
CC      CC      and for in vitro studies of differentiation of chondrocytes; and as
CC      CC      a target for anti-adhesive drugs or molecules in tendon, ligament,
CC      CC      skeletal muscle or other tissues where adhesion impairs the function
CC      CC      of the tissue (all claimed). 1Sa10 binding entities can be used to
CC      CC      determine the differentiation-state of cells during embryonic
CC      CC      development, angiogenesis or development of cancer, in pathological
CC      CC      conditions such as rheumatoid arthritis, osteoarthritis or cancer,
CC      CC      in tissue regeneration or in therapeutic and physiological repairation
CC      CC      of cartilage (claimed). A vaccine comprising the integrin
CC      CC      heterodimer or subunit alpha-10 is also claimed. 1Sa10
CC      CC      polynucleotides, vectors, host cells and methods of producing
CC      CC      recombinant 1Sa10 are also claimed.
XX      XX      Sequence 1132 AA;
SQ      SQ

```

Query Match 83.4%; Score 973; DB 21; Length 1132;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 973; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPVTHLFLPLVFLNGLCPFLNLDENHRLFGPPRAEFGYVLOHVGCGGRWMLVGA 60


```

Db      1 MELPVTHLFLPLVLTGLGCSFNLDEHHPRLPFGPPPEAEFGYSVLQHVGGQRMMLVGA 60
Qy      61 FWDGSGDRGDVYRCVPGAHNAPCAKGLGDYOLGNSSHPAVMHLMGMSLLETDDGGG 120
Db      61 FWDGSGDRGDVYRCVPGAHNAPCAKGLGDYOLGNSSHPAVMHLMGMSLLETDDGGG 120
Qy      121 FMACPLMSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYIP 180
Db      121 FMACPLMSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYIP 180
Qy      181 MSEVOTFLRLVIGKLFIDPEQIQVGLVOYGESPVHEMSLGDPRTEEYVRAAKNLSRREG 240
Db      181 MSEVOTFLRLVIGKLFIDPEQIQVGLVOYGESPVHEMSLGDPRTEEYVRAAKNLSRREG 240
Qy      241 RETKTAQAIMVACTGFSQSHGRPEARALVVVTDGSHDEELPALAKACEARVTRY 300
Db      241 RETKTAQAIMVACTGFSQSHGRPEARALVVVTDGSHDEELPALAKACEARVTRY 300
Qy      301 GIAVIGHYLRORDSSFLREIRITIASDPDERFFPNVTDAAALTDIVDALGDRIFGLSGS 360
Db      301 GIAVIGHYLRORDSSFLREIRITIASDPDERFFPNVTDAAALTDIVDALGDRIFGLSGS 360
Qy      361 HAENSSPGLSEMSQIGFSTHRLKDGILFGMVGA YDMGGSVLMLEGGHRLFPFRMLLEDEF 420
Db      361 HAENSSPGLSEMSQIGFSTHRLKDGILFGMVGA YDMGGSVLMLEGGHRLFPFRMLLEDEF 420
Qy      421 PPALONHAAYIGYSVSSMLRGRRFLSGAPFRHRRGVIAFOLKKGAVVAOSLOGE 480
Db      421 PPALONHAAYIGYSVSSMLRGRRFLSGAPFRHRRGVIAFOLKKGAVVAOSLOGE 480
Qy      481 QIGSYFSGBELCPLDTRDGGTTDVLIVAAPMFLGPONKKGTRVYVVLVQOQSILTLQSTLQ 540
Db      481 QIGSYFSGBELCPLDTRDGGTTDVLIVAAPMFLGPONKKGTRVYVVLVQOQSILTLQSTLQ 540
Qy      541 PEPPDADFSGFMAGLPDLNODGFADVAVGARLEDDHOGALVLYHGTSQGVRRPRAOIA 600
Db      541 PEPPDADFSGFMAGLPDLNODGFADVAVGARLEDDHOGALVLYHGTSQGVRRPRAOIA 600
Qy      601 AASMHAFSLYFGRSYDGRILDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Db      601 AASMHAFSLYFGRSYDGRILDGDDLVDVAVGAQGAAILLSRPIVHLTPSLEVTPOAIS 660
Qy      661 VVQDCCRRRGQAVCLTAALCFQVTSRTPGMRDHPFYMRFTASLDEWTRAGARAFDSGGQ 720
Db      661 VVQDCCRRRGQAVCLTAALCFQVTSRTPGMRDHPFYMRFTASLDEWTRAGARAFDSGGQ 720
Qy      721 RLSPPRLSLVGNWTCCEQLHFHVLDTSYLRLPVALTVPFALDNTTKPGPVLENGSPTSIQ 780
Db      721 RLSPPRLSLVGNWTCCEQLHFHVLDTSYLRLPVALTVPFALDNTTKPGPVLENGSPTSIQ 780
Qy      781 KLVPSKDCGPNCEVTDLVQNMNDIRGSRKAPFVVRGGRKVLVSTLLENKKNAYNT 840
Db      781 KLVPSKDCGPNCEVTDLVQNMNDIRGSRKAPFVVRGGRKVLVSTLLENKKNAYNT 840
Qy      841 SLISIIFFSRNLHLASITPQRESPIKYECAAPSAHAALCSVGHVFPOTGAKVTFLEFEFSC 900
Db      841 SLISIIFFSRNLHLASITPQRESPIKYECAAPSAHAALCSVGHVFPOTGAKVTFLEFEFSC 900
Qy      901 SSLLSQVFGKLTASSDSLERNGTLOENTASAYIQYEPHLLFSSSESTLHREYHAPYGTL 960
Db      901 SSLLSQVFGKLTASSDSLERNGTLOENTASAYIQYEPHLLFSSSESTLHREYHAPYGTL 960
Qy      961 FVGPGEPEFTTLR 973
Db      961 FVGPGEPEFTTLR 973

```

```

RESULT 4
AAB64658
ID      AAB64658 standard; Protein; 1152 AA.
XX
AC      AAB64658;
XX

```

```

Dt      22-MAR-2001 (first entry)
XX
XX      Human secreted protein BLAST search protein SEQ ID NO: 168.
DE
XX      Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX      antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX      cardiovascular; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX      cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX      neurological disease; infection; human; secreted protein.
XX
OS      Homo sapiens.
XX
XX      WO200077197-A1.
XX
XX      21-DEC-2000.
XX
XX      01-JUN-2000; 2000WO-US14934.
XX
XX      11-JUN-1999; 99US-0138599.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (ROSE/) ROSEN C A.
XX
XX      Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX      WPI; 2001-032312/04.
XX
XX      Disclosure: Page 547-551; 558pp; English.
XX
XX      The invention relates to the isolation of genes AAF3757-F32803 encoding
XX      the human secreted proteins AAB64549-B64594. The sequence is used as a
XX      query sequence for doing BLASTX searches to identify homologous
XX      sequences. The genes and proteins are useful for preventing,
XX      ameliorating or treating medical conditions, e.g. by protein or gene
XX      therapy. The genes are isolated from a range of human tissues disclosed
XX      in the specification. The nucleic acids, proteins, antibodies and
XX      (ant) agonists are useful in the diagnosis, treatment and prevention of:
XX      (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
XX      adrenal gland, bone, bone marrow, breast, gastrointestinal
XX      tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's
XX      disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX      thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX      rheumatoid arthritis and ulcerative colitis; (c) cardiovascular
XX      disorders such as myocardial ischaemia; (d) wound healing; (e)
XX      neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX      infectious diseases such as viral, bacterial, fungal and parasitic
XX      infections.
XX
XX      Sequence 1152 AA;
SQ

```

```

Query Match      36.6%; Score 427; DB 22; Length 1152;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 ELFPVTHLFLPLVLTGLGCSFNLDEHHPRLPFGPPPEAEFGYSVLQHVGGQRMMLVGA 61
Db      2 ELFPVTHLFLPLVLTGLGCSFNLDEHHPRLPFGPPPEAEFGYSVLQHVGGQRMMLVGA 61
Qy      62 WDGPSSDRGDVYRCVPGAHNAPCAKGLGDYOLGNSSHPAVMHLMGMSLLETDDGGG 121
Db      62 WDGPSSDRGDVYRCVPGAHNAPCAKGLGDYOLGNSSHPAVMHLMGMSLLETDDGGG 121
Qy      122 MACAPLMSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYIP 181
Db      122 MACAPLMSRACGSSVFSSGICARVDASFQPGSLAPTAQRCPTVMVYIVLDGNSIYIP 181
Qy      182 SEVOTFLRLVIGKLFIDPEQIQVGLVOYGESPVHEMSLGDPRTEEYVRAAKNLSRREG 241
Db      182 SEVOTFLRLVIGKLFIDPEQIQVGLVOYGESPVHEMSLGDPRTEEYVRAAKNLSRREG 241

```

QY 242 ETKTAQAIWVACTEGFSQSHGGRPEAPARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301
 DB 242 ETKTAQAIWVACTEGFSQSHGGRPEAPARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301
 QY 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESGH 361
 DB 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESGH 361
 QY 362 AENESSFGLIEMSOIGFSTRHLKDGIILFGWGAIVDMGGSVLMLEGGHRLFPFRMALEDEFP 421
 DB 362 AENESSFGLIEMSOIGFSTRHLKDGIILFGWGAIVDMGGSVLMLEGGHRLFPFRMALEDEFP 421
 QY 422 PALONHA 428
 DB 422 PALONHA 428

RESULT 5
 AAB64584
 ID AAB64584 standard; Protein; 1167 AA.
 XX
 AC AAB64584;
 XX
 DT 22-MAR-2001 (first entry)
 XX
 DE Human secreted protein #37.
 XX
 KM Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KM antiallergic; hepatocrotic; antidiabetic; antinflammatory; anticancer;
 KM vulnerrary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KM caridant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KM neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO20007197-A1.
 XX
 PD 21-DEC-2000.
 XX
 PF 01-JUN-2000; 2000MO-US14934.
 XX
 PR 11-JUN-1999; 99US-0138599.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis GA;
 XX
 DR WPI; 2001-032312/04.
 DR N-PSDB; AAF32793.
 PT Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 XX
 XX Claim 11; Page 496-500; 558bp; English.
 XX
 PS Sequences AAB64549-B64594 represent the amino acid sequences of 47
 CC human secreted proteins encoded by the genes AAF32757-F32803. The genes
 CC and proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from
 CC a range of human tissues disclosed in the specification. The nucleic
 CC acids, proteins, antibodies and (ant)agonists are useful in the
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC
 XX

SEQ Sequence 1167 AA;
 Query Match 36.6%; Score 427; DB 22; Length 1167;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 427; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ELFPVTHFLFLVFLVLTGLCSFPNLDENHRILFPGPPEAFSGYSVLQHVGGGQRMWLVGAP 61
 DB 2 ELFPVTHFLFLVFLVLTGLCSFPNLDENHRILFPGPPEAFSGYSVLQHVGGGQRMWLVGAP 61
 QY 62 WDGSFGRDGRGVYRCFVGGANAPCAKGLHDYOLGNSHPAVNMHLCMSLLETDDGCGF 121
 DB 62 WDGSFGRDGRGVYRCFVGGANAPCAKGLHDYOLGNSHPAVNMHLCMSLLETDDGCGF 121
 QY 122 MACAPLMSRACGSSVFSGICARVDASFQPGSLAFTQRCPTTMDVIVLDGNSITPW 181
 DB 122 MACAPLMSRACGSSVFSGICARVDASFQPGSLAFTQRCPTTMDVIVLDGNSITPW 181
 QY 182 SEVQTFRLRLVYKLFIDPEQIQVGLVOYGESPVHEWSLGDRTKEVVRAAKNLSRREG 241
 DB 182 SEVQTFRLRLVYKLFIDPEQIQVGLVOYGESPVHEWSLGDRTKEVVRAAKNLSRREG 241
 QY 242 ETKTAQAIWVACTEGFSQSHGGRPEAPARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301
 DB 242 ETKTAQAIWVACTEGFSQSHGGRPEAPARLLVVTVDGSHDEBELPALKACACAGRVTRYG 301
 QY 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESGH 361
 DB 302 IAVLGHYLRORDPSSFLREIRTIASDPDERFFFNVTDEALTDIYDALGDRIFGLESGH 361
 QY 362 AENESSFGLIEMSOIGFSTRHLKDGIILFGWGAIVDMGGSVLMLEGGHRLFPFRMALEDEFP 421
 DB 362 AENESSFGLIEMSOIGFSTRHLKDGIILFGWGAIVDMGGSVLMLEGGHRLFPFRMALEDEFP 421
 QY 422 PALONHA 428
 DB 422 PALONHA 428

RESULT 6
 AAU76853
 ID AAU76853 standard; Protein; 195 AA.
 XX
 AC AAU76853;
 XX
 DT 21-MAY-2002 (first entry)
 XX
 DE Human integrin alpha subunit Alpha 10 A domain.
 XX
 KM Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
 KM A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KM ischaemia-reperfusion injury; immune complex; parasitic disease;
 KM antinflammatory; vasotropic; antiparasitic; vulnerrary; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200209737-A1.
 XX
 PD 07-FEB-2002.
 XX
 PF 31-JUL-2001; 2001MO-US23957.
 XX
 PR 31-JUL-2000; 2000US-221950P.
 PR 11-JAN-2001; 2001US-0758493.
 PR 13-MAR-2001; 2001US-0805354.
 XX
 PA (GEMO) GEN HOSPITAL CORP.
 PA Arnaout AM, Li R, Xiong J;
 XX
 DR WPI; 2002-188687/24.
 XX
 PT Novel high affinity integrin polypeptide useful for treating restenosis
 XX

PT and parasitic diseases, comprises all or part of variant integrin alpha subunit A domain or variant integrin beta subunit A-like domain -

Example 2; Fig 5; 55pp; English.

CC The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD11b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC the human integrin alpha subunit Alpha 10 A domain.

XX Sequence 195 AA;

Query Match 16.7%; Score 195; DB 23; Length 195;
Best Local Similarity 100.0%; Pred. No. 6.3e-181;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSIGD 221
DB 1 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSIGD 60
OY 222 FRTKEEVVRAAKNLSRREGRETQTAQAIMVACTGEGFSOSHGREAPARLLVVTDSGSHD 281
DB 61 FRTKEEVVRAAKNLSRREGRETQTAQAIMVACTGEGFSOSHGREAPARLLVVTDSGSHD 120
OY 282 GEEIPALAKACEAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFNVTDDEA 341
DB 121 GEEIPALAKACEAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFNVTDDEA 180
OY 342 ALTDIVDALGDRIFG 356
DB 181 ALTDIVDALGDRIFG 195

RESULT 7

AAU76862 standard; Protein; 195 AA.

XX AAU76862;

DT 21-MAY-2002 (first entry)

XX Human integrin alpha subunit Alpha 10 variant A domain.

XX Human; integrin alpha subunit; A domain; Alpha 10; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
KW mutcin.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 193 /note= "Wild-type Ile substituted by any other amino
FT acid"

XX WO200209737-A1.

XX 07-FEB-2002.

XX 31-JUL-2001; 2001WO-US23957.

XX 31-JUL-2000; 2000US-221950P.
PR 11-JAN-2001; 2001US-0758493.
PR 13-MAR-2001; 2001US-0805354.

XX (GENO) GEN HOSPITAL CORP.

PI Arnaut AM, Li R, Xiong J;

XX WPI, 2002-188687/24.

XX Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PR subunit A domain or variant integrin beta subunit A-like domain -

XX Claim 53; Page -; 55pp; English.

XX The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD11b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC a human integrin alpha subunit Alpha 10 variant A domain.
CC Note: This variant sequence is not featured in the specification but has
CC been derived from the wild-type protein shown in AAU76863.

XX Sequence 195 AA;

Query Match 16.5%; Score 192; DB 23; Length 195;
Best Local Similarity 100.0%; Pred. No. 5.2e-178;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 162 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSIGD 221
DB 1 CPTVMDVVIYVLDGNSIYPMSEVQFLRLVGLKFLIDPEQIQVGLVQGESPVHEWSIGD 60
OY 222 FRTKEEVVRAAKNLSRREGRETQTAQAIMVACTGEGFSOSHGREAPARLLVVTDSGSHD 281
DB 61 FRTKEEVVRAAKNLSRREGRETQTAQAIMVACTGEGFSOSHGREAPARLLVVTDSGSHD 120
OY 282 GEEIPALAKACEAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFNVTDDEA 341
DB 121 GEEIPALAKACEAGRTVRYGIAVLGHYLRORDPSSFLREIRTIASDPDERFFNVTDDEA 180
OY 342 ALTDIVDALGDR 353
DB 181 ALTDIVDALGDR 192

RESULT 8

AAAY41735 standard; Protein; 124 AA.

XX AAAY41735;

DT 07-DEC-1999 (first entry)

XX Human PRO827 protein sequence.

XX Human; PRO: EST: expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.

XX Homo sapiens.

```

XX WO9946281-A2.
XX
XX 16-SEP-1999.
XX
XX 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079728.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 07-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.

PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085379.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086436.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.

XX
XX (GETH ) GENENTECH INC.
XX
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
XX WPI: 1999-551358/46.
XX N-PSDB; AA234185.
XX
XX New secreted and transmembrane polypeptides and their polynucleotides,
XX useful for treating blood coagulation disorders, cancers and cellular
XX adhesion disorders -
XX
XX Claim 12, Fig 139; 530pp; English.
XX
XX The present invention describes secreted and transmembrane polypeptides
XX and their polynucleotides. The nucleotide sequences are useful as
XX sources of probes, primers, for chromosome mapping, and for generation
XX of antisense sequences. They can also be used to create transgenic
XX animals. The proteins can be used to treat a variety of diseases and
XX disorders, depending on their function. Diseases that may be treated
XX include blood coagulation disorders, cancers and cellular adhesion
XX disorders. They may also be used to raise antibodies. AA233891 to
XX AA243338, and AA41685 to AA41774 represent polynucleotide and
XX polypeptide sequence given in the exemplification of the present
XX invention.
XX
XX Sequence 124 AA;
XX
SQ
Query Match 10.5%; Score 122; DB 20; Length 124;
Best Local Similarity 100.0%; Pred. No. 46e-110;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFRGPEAEFGYSVLQHVGGGRWMLVGA 60
DB 1 MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLFRGPEAEFGYSVLQHVGGGRWMLVGA 60
QY 61 PWDGSGGRGDYVRCPIVGGAHNAPCAKGHLGDYQLGNSSHPAVNMHLGMSLLETDGGG 120
DB 61 PWDGSGGRGDYVRCPIVGGAHNAPCAKGHLGDYQLGNSSHPAVNMHLGMSLLETDGGG 120
QY 121 FM 122
DB 121 FM 122

RESULT 9
AAB44291
ID AAB44291 standard; Protein: 124 AA.
XX AAB44291;
XX

```

DT 08-FEB-2001 (first entry)
XX Human PRO827 (UNQ468) protein sequence SEQ ID NO:346.
XX
XX Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
KM expressed sequence tag; detection; cancer.
OS Homo sapiens.
XX
XX WO200053756-A2.
PN
PD 14-SEP-2000.
XX
XF 18-FEB-2000; 2000WO-US04341.
XX
XX 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
PA
PA (GETH) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,
PI Ferrara N, Filvaroff E, Fong S, Gao W, Garber H, Gerritsen ME,
PI Goddard A, Goldowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ,
PI Kljavin LJ, Kuo SS, Napier MA, Pan J, Peoni NF, Roy MA,
PI Shelton DJ, Stewart TA, Tumas D, Williams PM, Wood WI;
XX MPI; 2000-611443/58.
DR N-PSDB; AAC78545.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
XX cellular activities -
XX
PS Claim 12; Fig 139; 636pp; English.

AAC78458 to AAC78599 represent polynucleotide and EST (expressed sequence tag) sequences which encode secreted or transmembrane PRO polypeptides. The PRO polynucleotides and polypeptides have cytostatic activity. The polynucleotides and polypeptides can be used for detecting the presence of PRO polypeptides in samples, for linking bioactive molecules to cells and for modulating biological activities of cells, using the polypeptides for specific targeting. The polypeptide targeting can be used to kill the target cells, e.g. for the treatment of cancers. The polypeptide pairs provide specific targeting of bioactive molecules to cells. AAC78600 to AAC78987 represent PCR primers and probes used in the isolation of the PRO polynucleotide sequences.

Sequence 124 AA;

```

Query Match 10.5%; Score 122; DB 21; Length 124;
Best Local Similarity 100.0%; Pred. No. 4.6e-110;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPVTITLFLPLVLTGICSPFNIDENHPRFLPDPPEAEFGYSYLQHVGGGRMMVLGA 60
Db 1 MELPVTITLFLPLVLTGICSPFNIDENHPRFLPDPPEAEFGYSYLQHVGGGRMMVLGA 60

```

Qy	61	PMWPGSGDRGDPVYRCPVGGAHNAACCAKGHIGDYOLGNSHPAVNMHLGMSLLETTDGDG	120
	61	PMWPGSGDRGDPVYRCPVGGAHNAACCAKGHIGDYOLGNSHPAVNMHLGMSLLETTDGDG	120
Db	61	PMWPGSGDRGDPVYRCPVGGAHNAACCAKGHIGDYOLGNSHPAVNMHLGMSLLETTDGDG	120
Qy	121	FM 122	
	121	FM 122	
Db	121	FM 122	

RESULT	10
AU29100	
ID	AAU29100 standard; Protein; 124 AA.
XX	
AC	AAU29100;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human PRO polypeptide sequence #77.
XX	
KW	PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KM	dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW	blood; chondrocyte cell; cell proliferation; colon; differentiation; colon;
KW	adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder
XX	
OS	Homo sapiens.
XX	
PN	WO200168848-A2.
XX	
PD	20-SEP-2001.
XX	
PF	28-FEB-2001; 2001WO-US066520.
XX	
PR	01-MAR-2000; 2000WO-US05601.
PR	02-MAR-2000; 2000WO-US05841.
PR	03-MAR-2000; 2000US-187202P.
PR	06-MAR-2000; 2000US-186968P.
PR	14-MAR-2000; 2000US-189320P.
PR	14-MAR-2000; 2000US-189328P.
PR	15-MAR-2000; 2000WO-US06884.
PR	21-MAR-2000; 2000US-190828P.
PR	21-MAR-2000; 2000US-191007P.
PR	21-MAR-2000; 2000US-191048P.
PR	21-MAR-2000; 2000US-191314P.
PR	28-MAR-2000; 2000US-192655P.
PR	29-MAR-2000; 2000US-193032P.
PR	29-MAR-2000; 2000US-193053P.
PR	30-MAR-2000; 2000WO-US08439.
PR	04-APR-2000; 2000US-194449P.
PR	04-APR-2000; 2000US-194647P.
PR	11-APR-2000; 2000US-195975P.
PR	11-APR-2000; 2000US-196000P.
PR	11-APR-2000; 2000US-196187P.
PR	11-APR-2000; 2000US-196690P.
PR	11-APR-2000; 2000US-196820P.
PR	18-APR-2000; 2000US-198121P.
PR	18-APR-2000; 2000US-198585P.
PR	23-APR-2000; 2000US-199377P.
PR	25-APR-2000; 2000US-199550P.
PR	25-APR-2000; 2000US-199654P.
PR	03-MAY-2000; 2000US-201516P.
PR	17-MAY-2000; 2000WO-US13705.
PR	22-MAY-2000; 2000WO-US14042.
PR	30-MAY-2000; 2000WO-US14841.
PR	02-JUN-2000; 2000WO-US15264.
PR	05-JUN-2000; 2000US-209832P.
PR	28-JUL-2000; 2000WO-US20710.
PR	22-AUG-2000; 2000US-0644848.
PR	24-AUG-2000; 2000WO-US23328.
PR	08-NOV-2000; 2000WO-US30952.
PR	01-DEC-2000; 2000WO-US32676.
PR	20-DEC-2000; 2000WO-US34956.
XX	
DA	(GETH) GENENTECH INC.

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX WPI; 2001-602746/68.
DR N-PSDB; AAS46001.
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX Claim 11; Fig 154; 774pp; English.
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
CC The PRO polypeptides and their associated nucleic acids can be used to
CC detect the presence of a tumour in a mammal by comparing the level of
CC expression of a PRO polypeptide in a test sample of cells from the animal
CC and a control sample of normal cells, whereby a higher level of
CC expression in the test sample indicates the presence of a tumour in the
CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
CC and rabbits but are preferably human. The polypeptides can be used to
CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
CC when contacted with it. A specific polypeptide can be used to stimulate
CC the proliferation or differentiation of chondrocyte cells. The PRO
CC proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX Sequence 124 AA;
SQ
Query Match 10.5%; Score 122; DB 22; Length 124;
Best Local Similarity 100.0%; Pred No. 4,6e-110;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELPVTHTLFLPLVLTGLCSFNNLDEHHPRLFPGPPEAFEGYSVLQHVGGGQRMVLVGA 60
DB 1 MELPVTHTLFLPLVLTGLCSFNNLDEHHPRLFPGPPEAFEGYSVLQHVGGGQRMVLVGA 60
QY 61 PWDGSGDRGDRGVYRCPVGANAPCAKHLGDYQIGNSSHAVNMHLMGSLLETDGDOG 120
DB 61 PWDGSGDRGDRGVYRCPVGANAPCAKHLGDYQIGNSSHAVNMHLMGSLLETDGDOG 120
QY 121 FM 122
DB 121 FM 122
RESULT 11
ABB95485
ID ABB95485 standard; Protein; 124 AA.
XX ABB95485;
AC ABB95485;
DT 19-JUL-2002 (first entry)
XX
DE Human angiogenesis related protein PRO827 SEQ ID NO: 126.
XX
XX Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiac; cytostatic; antiangiogenic; hypotensive; vulnery;
KW antiarteriosclerotic.
XX
OS Homo sapiens.
XX
XX WO200208284-A2.
XX 31-JAN-2002.
XX
XX 09-JUL-2001; 2001WO-US21735.
XX
XX 20-JUL-2000; 2000US-219556P.
PR

PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 15-SEP-2000; 2000US-000000P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242822P.
PR 08-NOV-2000; 2000US-0709238.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
PR 20-JUN-2001; 2001WO-US19692.
PR 28-JUN-2001; 2001WO-US00000.
XX
XX (GETH) GENENTECH INC.
PA (BAKE) BAKER K P.
PA (FERR) FERRARA N.
PA (GERB) GERBER H.
PA (GERR) GERRITSEN M E.
PA (GODD) GODDARD A.
PA (GODO) GODOWSKI P J.
PA (GURN) GURNEY A L.
PA (HILL) HILLAN K J.
PA (MARS) MARSTERS S A.
PA (PANC) PAN J.
PA (PAON) PAONI N F.
PA (STEP) STEPHAN J F.
PA (WATA) WATANABE C K.
PA (WILL) WILLIAMS P M.
PA (WOOD) WOOD W I.
XX
XX Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A,
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX WPI; 2002-171999/22.
XX N-PSDB; ABL95623.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 11; Fig 126; 567pp; English.
XX
XX The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangitis, tumour

CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
 CC healing. The present sequence is a PRO protein of the invention.
 XX
 SQ Sequence 124 AA;

Query Match 10.5%; Score 122; DB 23; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4.6e-110;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPVTHTLFLPLVLTGLCSFNLDEHPRLPFGPPPEAEFGYSVLQHVGGGQRMVLVGA 60
 DB 1 MELPVTHTLFLPLVLTGLCSFNLDEHPRLPFGPPPEAEFGYSVLQHVGGGQRMVLVGA 60
 QY 61 PWDGSGDRGDVYRCVPVGAHNAAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDCGG 120
 DB 61 PWDGSGDRGDVYRCVPVGAHNAAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDCGG 120
 QY 121 FM 122
 DB 121 FM 122

RESULT 12
 ABB84879
 ID ABB84879 standard; Protein; 124 AA.

AC ABB84879;

DT 16-MAY-2002 (first entry)

DE Human PRO827 protein sequence SEQ ID NO:126.

KW Human; angiogenesis; cardiac; cytoskeletal; antiangiogenic; hypotensive;
 KW vulnerability; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
 KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
 KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
 KW age-related macular degeneration; arterial restenosis; angina;
 KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
 KW lymphangitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
 KW wound healing; chromosome mapping; gene mapping.

OS Homo sapiens.

PN WO200200630-A2.

PD 03-JAN-2002.

FE 20-JUN-2001; 2001WO-US19692.

XX 23-JUN-2000; 2000US-213637P.
 PR 20-JUL-2000; 2000US-219556P.
 PR 25-JUL-2000; 2000US-220624P.
 PR 25-JUL-2000; 2000US-220664P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 02-AUG-2000; 2000US-222659P.
 PR 17-AUG-2000; 2000US-0643657.
 PR 23-AUG-2000; 2000WO-US23322.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 07-SEP-2000; 2000US-230378P.
 PR 18-SEP-2000; 2000US-0664610.
 PR 18-SEP-2000; 2000US-0665350.
 PR 24-OCT-2000; 2000US-242922P.
 PR 08-NOV-2000; 2000US-070923P.
 PR 08-NOV-2000; 2000WO-US10952.
 PR 10-NOV-2000; 2000WO-US30873.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000US-0747259.
 PR 22-JAN-2001; 2001US-0767609.
 PR 28-FEB-2001; 2001US-0796498.
 PR 28-FEB-2001; 2001WO-US06520.
 PR 01-MAR-2001; 2001WO-US06666.
 PR 09-MAR-2001; 2001US-0802706.

PR 14-MAR-2001; 2001US-0808689.
 PR 22-MAR-2001; 2001US-0816744.
 PR 05-APR-2001; 2001US-0828366.
 PR 10-MAY-2001; 2001US-0854208.
 PR 10-MAY-2001; 2001US-0854280.
 PR 25-MAY-2001; 2001US-0860028.
 PR 25-MAY-2001; 2001US-0866034.
 PR 25-MAY-2001; 2001WO-US17092.
 PR 30-MAY-2001; 2001US-0870574.
 PR 30-MAY-2001; 2001WO-US17443.
 PR 01-JUN-2001; 2001WO-US17800.

XX (GENTH) GENENTECH INC.

PA Baker KP, Ferrara N, Gerber H, Gertsen ME, Goddard A,
 PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF,
 PI Stephan WF, Watanabe CK, Williams PM, Wood WI, Ye W;

XX WPI; 2002-090516/12.
 DR N-PDDB; ABL88134.

XX One hundred and eighty seven nucleic acids encoding PRO polypeptides;
 PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
 PT infarction), endothelial or angiogenic disorders in a mammal -
 XX Claim 11; Fig 126; 565pp; English.

CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
 CC ABB85003. The PRO proteins and polynucleotides have cardiant, cytoskeletal,
 CC antiangiogenic, hypotensive, vulnerability and antiarteriosclerotic
 CC activities, and can be used in gene therapy. The PRO polynucleotides,
 CC proteins, agonists and antagonists are useful for treating or diagnosing
 CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
 CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
 CC degeneration, atherosclerosis, hypertension, arterial restenosis,
 CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
 CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
 CC carcinoma) and wound healing. The PRO polynucleotides have applications
 CC in molecular biology, including use as hybridisation probes, and in
 CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
 CC probes used in the exemplification of the present invention.

XX SQ Sequence 124 AA;

Query Match 10.5%; Score 122; DB 23; Length 124;
 Best Local Similarity 100.0%; Pred. No. 4.6e-110;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELPVTHTLFLPLVLTGLCSFNLDEHPRLPFGPPPEAEFGYSVLQHVGGGQRMVLVGA 60
 DB 1 MELPVTHTLFLPLVLTGLCSFNLDEHPRLPFGPPPEAEFGYSVLQHVGGGQRMVLVGA 60
 QY 61 PWDGSGDRGDVYRCVPVGAHNAAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDCGG 120
 DB 61 PWDGSGDRGDVYRCVPVGAHNAAPCAKGLGDYQLGNSSHPAVNMHLGMSLLETGDCGG 120
 QY 121 FM 122
 DB 121 FM 122

RESULT 13
 AAY32282
 ID AAY32282 standard; Protein; 303 AA.

AC AAY32282;

DT 15-FEB-2000 (first entry)

DE Mouse integrin subunit alpha-10 partial sequence.

KW Integrin alpha-10; I5a10; mouse; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;

Key	Location/Qualifiers
Misc-difference 15	/note= "encoded by TWG"
MO9951639-A1.	
14-OCT-1999.	
31-MAR-1999;	99WO-SE00544.
02-APR-1998;	98SE-0001164.
28-JAN-1999;	99SE-0000319.
(ACTI-) ACTIVE BIOTECH AB.	
Lundgren-Akerlund E;	
WPI: 2000-052639/04.	
N-PSDB; AAZ34635.	
New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation -	
Example 15; Fig 15; 90pp; English.	
This sequence represents a polypeptide encoded by an isolated partial genomic DNA (see AAZ34635) of murine integrin subunit alpha-10 (Isa10). The mouse genomic alpha-10 sequence was used to generate a targeting vector for knockout experiments. The invention relates to a recombinant or isolated integrin heterodimer comprising the alpha10 subunit in association with subunit beta (especially beta-1). The heterodimer and the subunit alpha-10 can be used as markers or targets of all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also be used for treating pathological conditions involving Isa10, such as: damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for detecting the formation of cartilage during embryonal development, physiological or therapeutic repair of cartilage, or detecting, regeneration of cartilage or chondrocytes during transplantation of cartilage or chondrocytes for selection and analysis or for sorting, isolating or purification of chondrocytes and for in vitro studies of differentiation of chondrocytes; and as a target for anti-adhesive drugs or molecules in tendon, ligament, skeletal muscle or other tissues where adhesion impairs the function of the tissue (all claimed). Binding entities, polynucleotides (PNS) or oligonucleotides (ONS) can be used for e.g. determining the differentiation-state of cells during embryonic development, angiogenesis or development of cancer, in pathological conditions such as rheumatoid arthritis, osteoarthritis or cancer, in tissue regeneration or in therapeutic and physiological repair of cartilage (claimed). The PNS and ONS can also be used for detecting pathological conditions such as atherosclerosis or inflammation (claimed). A vaccine comprising the integrin heterodimer or subunit alpha-10 is also claimed. Vectors, host cells and methods of producing recombinant Isa10 are provided.	
Sequence 303 AA;	
Query March 3 6%; Score 42; DB 21; Length 303;	
Best Local Similarity 100.0%; Pred. No. 7.2e-32;	
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0	

XX	RESULT 14
XX	AAV32244
ID	AAV32244 standard; Peptide; 22 AA.
XX	
AC	AAV32244;
XX	
DT	15-FEB-2000 (first entry)
XX	
DE	Human integrin subunit alpha-10 cytoplasmic domain peptide.
XX	
KW	Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis; osteoarthritis; osteoarthritis; cancer; atherosclerosis; inflammation; therapy; cartilage; chondrocyte; osteoblast; fibroblast; vaccine; marker.
KX	
OS	Homo sapiens.
XX	
PN	MO9951639-A1.
XX	
PD	14-OCT-1999.
XX	
PF	31-MAR-1999; 99NO-SF00544.
PR	'02-APR-1998; 98SE-0001164. 28-JAN-1999; 99SE-0000319.
PA	(ACTI-) ACTIVE BIOTECH AB.
XX	
P1	Lundgren-Akerlund E;
DR	WPI; 2000-052639/04.
PT	New isolated integrin subunit alpha-10, used as a marker or target molecule for cells during development, regeneration and pathological conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or inflammation -
PS	Claim 21; Page 53; 90pp; English.
XX	
CC	This sequence represents a fragment of novel human chondrocyte integrin subunit alpha-10 (ISa10, see AAV32242), corresponding to the C-terminal cytoplasmic domain of the protein. The invention relates to a recombinant or isolated integrin heterodimer comprising the alpha10 subunit in association with subunit beta (especially beta-1). The integrin heterodimer, or the subunit alpha-10, or a fragment of it such as the present sequence, can be used as a marker or target of all types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts. They can also be used: for treating pathological conditions involving ISa10, such as damage to cartilage, trauma, rheumatoid arthritis or osteoarthritis; for detecting the formation of cartilage during embryonal development, physiological or therapeutic repair of cartilage, or detecting regeneration of cartilage or chondrocytes during transplantation of cartilage or chondrocytes; for selection and analysis or for sorting, isolating or purification of chondrocytes and for in vitro studies of differentiation of chondrocytes; and as a target for anti-adhesive drugs or molecules in tendon, ligament, or skeletal muscle or other tissues where adhesion impairs the function of the tissue (all claimed). ISa10 binding entities can be used to determine the differentiation-state of cells during embryonic development, angiogenesis or development of cancer, in pathological conditions such as rheumatoid arthritis, osteoarthritis or cancer, in tissue regeneration or in therapeutic and physiological repair of cartilage (reclaimed). A vaccine comprising the integrin heterodimer or subunit alpha-10 is also claimed. ISa10 polynucleotides, vectors, host cells and methods of producing recombinant ISa10 are also claimed.
SO	Sequence 22 AA;
Query Match	1.9%; Score 22; DB 21; Length 22;
Best Local Similarity	100.0%; Pred. No. 2.le-13;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1146 KLGFFAHKKIPPEEKREKLEQ 1167
| | | | | | | | | | | | | | | | | | | | | |
DB 1 KLGFFAHKKIPPEEKREKLEQ 22

RESULT 15
AAV32245
ID AAV32245 standard; Peptide; 18 AA.
XX
XX AAV32245;
AC
XX
XX 15-FEB-2000 (first entry)
DT
DE Bovine integrin subunit alpha-10 tryptic peptide 1.
XX
XX Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
KM osteoarthritis; osteoarthritis; cancer; atherosclerosis;
KM inflammation; therapy; cartilage; chondrocyte; osteoblast;
KM fibroblast; vaccine; marker.
XX
XX Bos taurus.
OS
XX W09951639-A1.
PN
XX 14-OCT-1999.
PD
XX 31-MAR-1999; 99MO-SE00544.
PF
XX 02-APR-1998; 98SE-0001164.
PR
XX 28-JAN-1999; 99SE-0000319.
PR
XX (ACTI-) ACTIVE BIOTECH AB.
PA
XX
XX Lundgren-Akerlund E;
PI
XX WPI; 2000-052639/04.
DR
XX
XX New isolated integrin subunit alpha-10, used as a marker or target
PT molecule for cells during development, regeneration and pathological
PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
PT inflammation -
PT
XX
XX Example 2; Fig 2; 90pp; English.
PS
XX
XX This sequence represents tryptic peptide 1 of integrin subunit
CC alpha-10 (ISa10) isolated from bovine chondrocytes. 6 tryptic
CC peptides (see AAV32245-50) were obtained. PCR primers (see
CC AAZ34721-24) based on peptides 1 and 2 were used to obtain bovine
CC ISa10 CDNA fragments. These were used to screen a human
CC articular chondrocyte cDNA library in order to isolate human ISa10
CC cDNA (see AAZ34719). The invention relates to a recombinant or
CC isolated integrin heterodimer comprising the alpha-10 subunit in
CC association with subunit beta (especially beta-1). The heterodimer
CC or the subunit alpha-10 can be used as a marker or target of all
CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
CC They can also be used as active ingredients in pharmaceutical
CC compositions and vaccines.
CC
XX
XX Sequence 18 AA;
SQ

Query Match 1.2%; Score 14; DB 21; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 927 NTAQTSAYIQEYEPH 940
| | | | | | | | | | | | | | | | | | | | | |
DB 2 NTAQTSAYIQEYEPH 15

RESULT 16
AAU87675

ID AAU87675 standard; Protein; 103 AA.
XX
XX AAU87675;
AC
XX
XX 05-JUN-2002 (first entry)
DT
XX
DE Novel central nervous system protein #585.
XX
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KM hyperproliferative disorder; neoplasm; cardiovascular disorder;
KM cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;
KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KM adenocarcinoma; reproductive system disorder; testicular feminisation;
KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KM respiratory disorder; renal disorder; kidney failure; blood disorder;
KM myocardial infarction; wound healing; cell proliferation; skin aging;
KM food additive; food preservative; gene therapy.
XX
XX Homo sapiens.
OS
XX W020015318-A2.
PN
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01332.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR
XX 04-FEB-2000; 2000US-0180628.
PR
XX 24-FEB-2000; 2000US-0184664.
PR
XX 02-MAR-2000; 2000US-0186350.
PR
XX 16-MAR-2000; 2000US-0189874.
PR
XX 17-MAR-2000; 2000US-0190076.
PR
XX 18-APR-2000; 2000US-0198123.
PR
XX 19-MAY-2000; 2000US-0205515.
PR
XX 07-JUN-2000; 2000US-0209467.
PR
XX 28-JUN-2000; 2000US-0214886.
PR
XX 30-JUN-2000; 2000US-0215135.
PR
XX 07-JUL-2000; 2000US-0216647.
PR
XX 07-JUL-2000; 2000US-0216880.
PR
XX 11-JUL-2000; 2000US-0217487.
PR
XX 11-JUL-2000; 2000US-0217496.
PR
XX 14-JUL-2000; 2000US-0218290.
PR
XX 26-JUL-2000; 2000US-0220963.
PR
XX 26-JUL-2000; 2000US-0220964.
PR
XX 14-AUG-2000; 2000US-0224519.
PR
XX 14-AUG-2000; 2000US-0225213.
PR
XX 14-AUG-2000; 2000US-0225214.
PR
XX 14-AUG-2000; 2000US-0225266.
PR
XX 14-AUG-2000; 2000US-0225267.
PR
XX 14-AUG-2000; 2000US-0225268.
PR
XX 14-AUG-2000; 2000US-0225270.
PR
XX 14-AUG-2000; 2000US-0225447.
PR
XX 14-AUG-2000; 2000US-0225757.
PR
XX 14-AUG-2000; 2000US-0225758.
PR
XX 14-AUG-2000; 2000US-0225759.
PR
XX 18-AUG-2000; 2000US-0226279.
PR
XX 22-AUG-2000; 2000US-0226681.
PR
XX 22-AUG-2000; 2000US-0226868.
PR
XX 22-AUG-2000; 2000US-0227182.
PR
XX 23-AUG-2000; 2000US-0227009.
PR
XX 30-AUG-2000; 2000US-0228924.
PR
XX 01-SEP-2000; 2000US-0229287.
PR
XX 01-SEP-2000; 2000US-0229343.
PR
XX 01-SEP-2000; 2000US-0229344.
PR
XX 01-SEP-2000; 2000US-0229345.
PR
XX 05-SEP-2000; 2000US-0229509.
PR
XX 05-SEP-2000; 2000US-0229513.
PR
XX 06-SEP-2000; 2000US-0230437.
PR
XX 06-SEP-2000; 2000US-0230438.
PR
XX 08-SEP-2000; 2000US-0231242.
PR
XX 08-SEP-2000; 2000US-0231243.

[illegible]

	PR	17-NOV-2000;	2000US-0249265.
	PR	17-NOV-2000;	2000US-0249297.
	FR	17-NOV-2000;	2000US-0249299.
	FR	17-NOV-2000;	2000US-0249300.
	PR	01-DEC-2000;	2000US-0250160.
	PR	01-DEC-2000;	2000US-0250391.
	PR	05-DEC-2000;	2000US-0251030.
	PR	05-DEC-2000;	2000US-0251988.
	PR	05-DEC-2000;	2000US-0256719.
	PR	06-DEC-2000;	2000US-0251479.
	PR	08-DEC-2000;	2000US-0251856.
	PR	08-DEC-2000;	2000US-0251868.
	PR	08-DEC-2000;	2000US-0251869.
	PR	08-DEC-2000;	2000US-0251989.
	PR	08-DEC-2000;	2000US-0251990.
	PR	11-DEC-2000;	2000US-0254097.
	FR	05-JAN-2001;	2001US-0259678.
	PA	(HUMA-) HUMAN GENOME SCI INC.	
	PJ	Rosen CA,	Barrash SC, Ruben SM;
	XX		
	DR	WPI:	2001-581633/65.
	PT	N-PSDB; ABRK44005.	
	PT	New isolated nucleic acid encoding a protein for diagnosing, preventing, treating or ameliorating medical conditions and used as food additives or preservatives -	
	PS	Claim 9; SEQ ID No 1193; 837pp; English.	
	CC	The invention describes an isolated nucleic acid molecule (I) encoding a novel central nervous system protein. (II) and polypeptides (III) encoded by (I), are used to treat a medical condition and in diagnosis of a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease and amyotrophic lateral sclerosis, infections caused by bacteria, viruses e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders e.g. corneal infection, gastrointestinal disorders e.g. dysphagia, adenocarcinomas and irritable bowel syndrome, reproductive system disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes mellitus and pituitary dwarfism, cancers and disorders at the cellular level e.g. leukemia, disorders involving neovascularization e.g. malignancies, respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g. acute kidney failure and blood related disorders e.g. myocardial infarction. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotherapies. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein,	
	OY	Query Match	1.1k; Score 13; DB 22; Length 103; Best Local Similarity 100.0%; Pred. NO. 0.00047; Indels 0; Gaps 0. Matches 13; Conservative 0; Mismatches 0;
	Dd	169 VIVLDGNSNITPM 181 15 VIVLDGSNTIPM 27	
	RESULT 17		
	ID	AU19822 standard; Protein; 103 AA.	
	AC	AU19822;	
	Tt	06-DEC-2001 (first entry)	
	XE	Human novel extracellular matrix protein, Seq ID No 472.	

XX Human; secreted extracellular matrix protein; immunomodulatory;
KW Anti-HIV; antianemic; antithematic; antisclerotic; cardiac; vascular;
KW cerebroprotective; chyrolytic; antimicrobial; ophthalmic; cytostatic;
KW antiallergic; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antisense; food additive.
OS Homo sapiens.
PN WO20015368-A1.
XX
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01348.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180626.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 14-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231480.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.

PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

XX Claim 11; SEQ ID NO 472; 235bp + Sequence Listing; English.
PS
XX
CC The invention relates to novel genes (AB06521-AB06785) and proteins
CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=99909764870.
CC
SQ Sequence 103 AA;
Query Match 1.1%; Score 13; DB 23; Length 103;
Best Local Similarity 100.0%; Pred. No. 0.00047;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 169 VVLDGNSNIYPW 181
Db 15 VVLDGNSNIYPW 27
RESULT 19
AAU19634
ID AAU19634 standard; Protein; 148 AA.
XX
AC AAU19634;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human novel extracellular matrix protein, Seq ID No 284.
XX
KW Human; secreted extracellular matrix protein; immunomodulatory;
KW Anti-HIV; antianemic; antitubercular; antisclerotic; cardiac; vascular;
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW antialzheimer; immune/autoimmune disease; HIV infection; anaemia;
KW human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW cancer; hyperproliferative disorder; breast neoplasm; melanoma;
KW Sezary syndrome; Gaucher's disease; neurological diseases;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW cardiac arrest; tachycardia; angina; infection; corneal infections;
KW wound healing; immunogen; gene therapy; antilease; food additive.
XX
OS Homo sapiens.
XX
PN WO200155368-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01348.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239933.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.

PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225265.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0231080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0234000.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250191.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251968.
 PR 05-DEC-2000; 2000US-0251969.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 PI
 XX
 DR WPI; 2001-465572/50.
 DR N-PSDB; AAS31365.
 XX
 XX Nucleic acid molecules encoding human secreted extracellular matrix
 PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
 PT Alzheimer's and Parkinson's diseases and cancers -
 XX
 XX
 XX Claim 11; SEQ ID No 444; 577bp; English.
 XX
 XX The invention relates to isolated nucleic acid molecules encoding
 CC novel human secreted extracellular matrix proteins (SPs). The
 CC polynucleotides and proteins are used to prevent, treat a medical
 CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
 CC chickens or sheep. For example, disorders associated with decreased
 CC expression of SPs. The SP polynucleotide or a vector expressing them may
 CC be administered to treat diseases by gene therapy. Antisense molecules
 CC may be administered to down regulate expression of SPs by binding with
 CC the cells own genes and preventing their expression. The polynucleotides
 CC may also be used as DNA probes in diagnostic assays. The SPs may also be
 CC used as antigens to produce antibodies and to identify modulators
 CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
 CC antagonists may also be used to down regulate expression and activity of
 CC SP and as diagnostic agents for detecting the presence of SPs in samples.

CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
 CC arrest, tachycardia and angina), infections caused by bacteria, viruses
 CC and fungi and ocular disorders (e.g. corneal infections). Other uses
 CC include wound healing, maintenance of organs before transplantation,

Query Match 1.1%; Score 13; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 0.00064;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
 |||||
 Db 49 VIVLDGNSIYPM 61

RESULT 21

ID ABP47854 standard; Protein; 148 AA.

XX AC ABP47854;

XX DT 23-AUG-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 284.

KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
 KW antirheumatic; hepatoprotective; cerebroprotective; antineuroinflammatory;
 KW antiallergic; antidiabetic; antituber; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotoxic; gene therapy; vaccine.

XX OS Homo sapiens.

XX PN US2002042386-A1.

XX PD 11-APR-2002.

XX PF 17-JAN-2001; 2001US-0764870.

XX PR 31-JAN-2000; 2000US-179065P.

XX PR 04-FEB-2000; 2000US-180628P.

XX PR 28-JUN-2000; 2000US-214886P.

XX PR 07-JUL-2000; 2000US-216647P.

XX PR 11-JUL-2000; 2000US-216880P.

XX PR 11-JUL-2000; 2000US-217487P.

XX PR 14-JUL-2000; 2000US-218290P.

XX PR 26-JUL-2000; 2000US-220963P.

XX PR 26-JUL-2000; 2000US-220964P.

XX PR 14-AUG-2000; 2000US-224518P.

XX PR 14-AUG-2000; 2000US-224519P.

XX PR 14-AUG-2000; 2000US-225267P.

XX PR 14-AUG-2000; 2000US-225268P.

XX PR 14-AUG-2000; 2000US-225270P.

XX PR 14-AUG-2000; 2000US-225447P.

XX PR 14-AUG-2000; 2000US-225758P.

XX PR 22-AUG-2000; 2000US-226868P.

XX PR 30-AUG-2000; 2000US-228934P.

XX PR 01-SEP-2000; 2000US-229267P.

XX PR 01-SEP-2000; 2000US-229343P.

XX PR 01-SEP-2000; 2000US-229344P.

XX PR 01-SEP-2000; 2000US-229345P.

XX PR 05-SEP-2000; 2000US-229509P.

XX PR 08-SEP-2000; 2000US-229513P.

XX PR 21-SEP-2000; 2000US-231413P.

XX PR 21-SEP-2000; 2000US-234223P.

PR 21-SEP-2000; 2000US-234274P.
 PR 25-SEP-2000; 2000US-234997P.
 PR 27-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236368P.
 PR 29-SEP-2000; 2000US-236369P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 02-OCT-2000; 2000US-237040P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 01-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.

XX (ROSE/) ROSEN C A.
 XX (RUBE/) RUBEN S M.
 XX (BARA/) BARASH S C.

PI Rosen CA, Ruben SM, Barash SC;

XX WPI, 2002-470713/50.

XX N-PSDB; AB066529.

PT New nucleic acid encoding human proteins, useful for diagnosis,
 PT treatment and prevention of e.g. osteoporosis, also related
 PT polypeptides and antibodies

XX Claim 11; SEQ ID NO 284; 215pp + Sequence Listing; English.

XX The invention relates to novel genes (AB066521-AB066785) and proteins
 CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?docid=99909764870.

XX Sequence 148 AA;

Query Match 1.1%; Score 13; DB 23; Length 148;
 Best Local Similarity 100.0%; Pred. No. 0.00064;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
 |||||
 Db 49 VIVLDGNSIYPM 61

RESULT 22

ID ABP48014 standard; Protein; 148 AA.

XX AC ABP48014;

XX	DT	.23-AUG-2002	(first entry)
XX	DE	Human polypeptide SEQ ID NO 444.	
XX	KW	Human; nocitropic; neuroprotective; cytostatic; dermatological; antiviral;	
KW	KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnereary	
KW	KW	antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;	
KW	KW	antihaemetic; hepatotropic; cerebroprotective; antiinflammatory;	
KW	KW	antiallergic; antidiabetic; antifurcer; anticonvulsant; antifungal;	
KW	KW	antiparasitic; cardiac; immune disorder; cardiovascular disorder;	
XX	KW	neurological disease; infection; nephrotropic; gene therapy; vaccine.	
OS		Homo sapiens.	
XX	PN	US2002042386-A1.	
XX	PD	11-APR-2002.	
XX	PE	17-JAN-2001; 2001US-0764870.	
XX	PR	31-JAN-2000; 2000US-179065P.	
PR	PR	04-FEB-2000; 2000US-180628P.	
PR	PR	28-JUN-2000; 2000US-214886P.	
PR	PR	07-JUL-2000; 2000US-216647P.	
PR	PR	07-JUL-2000; 2000US-216880P.	
PR	PR	11-JUL-2000; 2000US-217487P.	
PR	PR	11-JUL-2000; 2000US-217496P.	
PR	PR	14-JUL-2000; 2000US-218290P.	
PR	PR	26-JUL-2000; 2000US-220963P.	
PR	PR	26-JUL-2000; 2000US-220964P.	
PR	PR	14-AUG-2000; 2000US-224518P.	
PR	PR	14-AUG-2000; 2000US-224519P.	
PR	PR	14-AUG-2000; 2000US-225267P.	
PR	PR	14-AUG-2000; 2000US-225268P.	
PR	PR	14-AUG-2000; 2000US-225270P.	
PR	PR	14-AUG-2000; 2000US-225447P.	
PR	PR	14-AUG-2000; 2000US-225757P.	
PR	PR	14-AUG-2000; 2000US-225758P.	
PR	PR	22-AUG-2000; 2000US-226688P.	
PR	PR	30-AUG-2000; 2000US-228924P.	
PR	PR	01-SEP-2000; 2000US-229287P.	
PR	PR	01-SEP-2000; 2000US-229343P.	
PR	PR	01-SEP-2000; 2000US-229344P.	
PR	PR	01-SEP-2000; 2000US-229345P.	
PR	PR	05-SEP-2000; 2000US-229509P.	
PR	PR	05-SEP-2000; 2000US-229513P.	
PR	PR	08-SEP-2000; 2000US-231413P.	
PR	PR	21-SEP-2000; 2000US-231423P.	
PR	PR	21-SEP-2000; 2000US-234274P.	
PR	PR	25-SEP-2000; 2000US-234977P.	
PR	PR	27-SEP-2000; 2000US-235834P.	
PR	PR	29-SEP-2000; 2000US-236327P.	
PR	PR	29-SEP-2000; 2000US-236367P.	
PR	PR	29-SEP-2000; 2000US-236368P.	
PR	PR	29-SEP-2000; 2000US-236369P.	
PR	PR	29-SEP-2000; 2000US-236370P.	
PR	PR	02-OCT-2000; 2000US-236802P.	
PR	PR	02-OCT-2000; 2000US-237037P.	
PR	PR	02-OCT-2000; 2000US-237038P.	
PR	PR	02-OCT-2000; 2000US-237039P.	
PR	PR	13-OCT-2000; 2000US-237040P.	
PR	PR	13-OCT-2000; 2000US-239353P.	
PR	PR	20-OCT-2000; 2000US-240960P.	
PR	PR	20-OCT-2000; 2000US-241785P.	
PR	PR	20-OCT-2000; 2000US-241809P.	
PR	PR	01-NOV-2000; 2000US-244617P.	
PR	PR	17-NOV-2000; 2000US-249299P.	
PR	PR	08-DEC-2000; 2000US-251856P.	
PR	PR	08-DEC-2000; 2000US-251866P.	
PR	PR	08-DEC-2000; 2000US-251869P.	
PA	XA	(ROSE/) ROSEN C A.	

PA	(RUBE)/ RUBEN S M.
PA	(BARA)/ BARASH S C.
PI	Rosen CA, Ruben SM, Barash SC;
DR	WPI: 2002-470713/50.
DR	N-PSDB; ABQ66689.
XX	
PT	New nucleic acid encoding human proteins, useful for diagnosis,
PT	treatment and prevention of e.g. osteoporosis, also related
PT	polypeptides and antibodies -
PS	Claim 11; SEQ ID NO 444; 235pp + Sequence Listing; English.
XX	
XX	The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
CC	(ABP47846-ABP48110) useful for preventing, treating or ameliorating
CC	medical conditions e.g. by protein or gene therapy. The genes are
CC	isolated from a range of human tissues disclosed in the specification.
CC	The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC	and parasitic infections.
CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from USPTO at seqdata.uspto.gov/sequence.html?docid=999909764870.
XX	
XX	
SQ	Sequence 148 AA;
Query Match:	1.1%; Score 13; DB 23; Length 148;
Best Local Similarity	100.0%; Pred. No. 0.00064;
Matches 13; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	169 VIVLDGSNSIYPM 181
Db	49 VIVLDGSNSIYPM 61
RESULT 23	
AAU76854	
ID	AAU76854 standard; Protein: 193 AA.
XX	
AC	AAU76854;
XX	
DT	21-MAY-2002 (first entry)
DE	
XX	Human integrin alpha subunit Alpha 11 A domain.
KW	Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;
KW	A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW	ischaemia-reperfusion injury; immune complex; parasitic disease;
KW	antiinflammatory; vasotropic; antiparasitic; vulnerability; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	WO200209737-A1.
XX	
PD	07-FEB-2002.
XX	
PF	31-JUL-2001; 2001WO-US23957.
XX	
PR	31-JUL-2000; 2000US-221950P.
PR	11-JAN-2001; 2001US-0758493.
PR	13-MAR-2001; 2001US-0805354.
XX	
PA	(GEHO) GEN HOSPITAL CORP.
XX	

PI Arnaout AM, Li R, Xiong J;
XX
DR WPI; 2002-186687/24.
XX
XX Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain
XX
XX Example 2; Fig 5; 55pp; English.
PS
CC The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD1b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC the human integrin alpha subunit Alpha 11 A domain.
XX
SQ Sequence 193 AA;
XX
Query Match 1.1%; Score 13; DB 23; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 169 VIVLDGNSNIYPW 181
DB 8 VIVLDGNSNIYPW 20
RESULT 24
AAU76863
ID AAU76863 standard; Protein; 193 AA.
XX
AC AAU76863;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human integrin alpha subunit Alpha 11 variant A domain.
XX
KW Human; integrin alpha subunit; A domain; Alpha 11; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
KW mutuin.
XX
OS Homo sapiens.
OS Synthetic.
OS
FH Key Location/Qualifiers
FT Misc-difference 191
FT /note= "wild-type Ile substituted by any other amino
FT acid"
FT
FT
XX
XX WO200209737-A1.
XX
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US23957.
XX
XX 31-JUL-2000; 2000US-221950P.
XX 11-JAN-2001; 2001US-0758493.
XX 13-MAR-2001; 2001US-0805354.
XX
XX (GEHO) GEN HOSPITAL CORP.
XX
XX

PI Arnaout AM, Li R, Xiong J;
XX
DR WPI; 2002-186687/24.
XX
XX Novel high affinity integrin polypeptide useful for treating restenosis
PT and parasitic diseases, comprises all or part of variant integrin alpha
PT subunit A domain or variant integrin beta subunit A-like domain
XX
XX Claim 57; Page -; 55pp; English.
PS
CC The invention relates to a high affinity integrin polypeptide comprising
CC all or part of a variant integrin alpha subunit A domain or a variant
CC integrin beta subunit A-like domain. The polypeptide, preferably the
CC CD1b alpha subunit A domain, where I at residue 332 has been replaced by
CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
CC V at residue 315 and A at residue 320 have been replaced by C, is useful
CC for determining if a test compound is a candidate compound for binding to
CC CD1b or for treating an inflammatory disorder, by contacting a test
CC compound with the polypeptide and determining if the test compound binds
CC to the polypeptide. The integrin subunits are useful for reducing
CC skeletal muscle injury, for treating disorders caused by
CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
CC diseases, to purify variant integrin polypeptide ligands and as bait
CC proteins in two-hybrid or three-hybrid assays. This sequence represents
CC a human integrin alpha subunit Alpha 11 variant A domain.
CC Note: This variant sequence is not featured in the specification but has
CC been derived from the wild-type protein shown in AAU76854.
XX
SQ Sequence 193 AA;
XX
Query Match 1.1%; Score 13; DB 23; Length 193;
Best Local Similarity 100.0%; Pred. No. 0.00081;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 169 VIVLDGNSNIYPW 181
DB 8 VIVLDGNSNIYPW 20
RESULT 25
AAU76851
ID AAU76851 standard; Protein; 195 AA.
XX
AC AAU76851;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human integrin alpha subunit Alpha 1 (CD49a) A domain.
XX
KW Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;
KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
KW ischaemia-reperfusion injury; immune complex; parasitic disease; Alpha 1;
KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy.
XX
OS Homo sapiens.
OS
FH Key Location/Qualifiers
FT Misc-difference 191
FT /note= "wild-type Ile substituted by any other amino
FT acid"
FT
FT
XX
XX WO200209737-A1.
XX
XX 07-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US23957.
XX
XX 31-JUL-2000; 2000US-221950P.
XX 11-JAN-2001; 2001US-0758493.
XX 13-MAR-2001; 2001US-0805354.
XX
XX (GEHO) GEN HOSPITAL CORP.
XX
XX Arnaout AM, Li R, Xiong J;
XX
XX WPI; 2002-186687/24.
XX
XX Novel high affinity integrin polypeptide useful for treating restenosis
PT

PT and parasitic diseases, comprises all or part of variant integrin alpha
 subunit A domain or variant integrin beta subunit A-like domain -
 XX
 PS Example 2; Fig 5; 55pp; English.

CC The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD11b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC the human integrin alpha subunit Alpha 1 (CD49a) A domain.
 CC
 XX

Sequence 195 AA;

Query Match 1.1%; Score 13; DB 23; Length 195;
 Best Local Similarity 100.0%; Pred. No. 0.00082;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSNIYPW 181
 |||||
 DB 8 VIVLDGNSNIYPW 20

RESULT 26

AAU76860
 ID AAU76860 standard; Protein; 195 AA.

AC AAU76860;

DT 21-MAY-2002 (first entry)

DE Human integrin alpha subunit Alpha 1 (CD49a) variant A domain.

XX Human; integrin alpha subunit; A domain; CD49a; integrin beta subunit;
 KW A-like domain; inflammatory disorder; skeletal muscle injury; restenosis;
 KW ischaemia-reperfusion injury; immune complex; parasitic disease; mutant;
 KW antiinflammatory; vasotropic; antiparasitic; vulnerary; gene therapy;
 KM mutin; Alpha 1.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 193 /note="Wild-type Ile substituted by any other amino
 FT acid"
 FT
 FT

PN WO200209737-A1.

PD 07-FEB-2002.

PF 31-JUL-2001; 2001WO-US23957.

PR 31-JUL-2000; 2000US-221950P.

PR 11-JAN-2001; 2001US-0758493.

PR 13-MAR-2001; 2001US-0805354.

PA (GEHO) GEN HOSPITAL CORP.

PI Arnaud AM, Li R, Xiong J;

PI WPI; 2002-188687/24.

PT Novel high affinity integrin polypeptide useful for treating restenosis

PT and parasitic diseases, comprises all or part of variant integrin alpha
 subunit A domain or variant integrin beta subunit A-like domain -
 XX
 PS Claim 45; Page -; 55pp; English.

CC The invention relates to a high affinity integrin polypeptide comprising
 CC all or part of a variant integrin alpha subunit A domain or a variant
 CC integrin beta subunit A-like domain. The polypeptide, preferably the
 CC CD11b alpha subunit A domain, where I at residue 332 has been replaced by
 CC G or A, F at residue 313 and A at residue 320 have been replaced by C, or
 CC V at residue 315 and A at residue 320 have been replaced by C, is useful
 CC for determining if a test compound is a candidate compound for binding to
 CC CD11b or for treating an inflammatory disorder, by contacting a test
 CC compound with the polypeptide and determining if the test compound binds
 CC to the polypeptide. The integrin subunits are useful for reducing
 CC skeletal muscle injury, for treating disorders caused by
 CC ischaemia-reperfusion injury, immune complexes, restenosis and parasitic
 CC diseases, to purify variant integrin polypeptide ligands and as bait
 CC proteins in two-hybrid or three-hybrid assays. This sequence represents
 CC a human integrin alpha subunit Alpha 1 (CD49a) variant A domain.
 CC Note: This variant sequence is not featured in the specification but has
 CC been derived from the wild-type protein shown in AAU76851.
 CC
 XX

Sequence 195 AA;

Query Match 1.1%; Score 13; DB 23; Length 195;
 Best Local Similarity 100.0%; Pred. No. 0.00082;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSNIYPW 181
 |||||
 DB 8 VIVLDGNSNIYPW 20

RESULT 27

AAB50041
 ID AAB50041 standard; Protein; 214 AA.

AC AAB50041;

DT 14-MAR-2001 (first entry)

DE Rat alpha1 integrin domain.

XX Rat; alpha1 integrin domain; alphasbeta1 function blocking antibody;
 DE inflammatory disorder; rheumatoid arthritis; skin condition; asthma;
 KW bronchitis; headache; antipyretic; fever; gastrointestinal;
 KW vascular disease; autoimmune; respiratory distress syndrome;
 KW endotoxin shock syndrome; atherosclerosis.
 XX
 OS Rattus sp.

XX WO200072881-A1.
 PN 07-DEC-2000.

PF 01-JUN-2000; 2000WO-US15004.

PR 01-JUN-1999; 99US-0137038.

PR 29-FEB-2000; 2000US-0185336.

PA (BIOJ) BIOGEN INC.

PI De Fougereolles A, Gotwals P, Lobb R, Koteliensky V;

PI WPI; 2001-061448/07.

PT Use of blocking monoclonal antibody capable of binding to an epitope of
 PT VLA-1 for treating inflammatory disorders, in particular arthritis -
 XX
 PS Example 14; Fig 14; 60pp; English.

PT The present invention relates to the use of an alphabeta1 function

KW Immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
 KW multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
 KW Crohn's disease; nephritis; hyperproliferative disorder;
 KW cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
 KW melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
 OS Homo sapiens.
 PN WO20029435-A1.
 PD 25-MAY-2000.
 PF 27-OCT-1999; 99WO-US25031.
 PR 28-OCT-1998; 98US-0105971.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,
 PI Greene JM;
 DR WPI; 2000-387742/33.
 XX
 PT Isolated nucleic acid molecules encoding human secreted proteins are
 PT used for the prevention, amelioration and treatment of autoimmune,
 PT inflammatory, hyperproliferative and cardiovascular disorders, cancer,
 PT wounds, and infectious diseases -
 PS Claim 1; Page 678-682; 803pp; English.
 XX
 CC The present invention relates to 12 secreted human proteins and the
 CC nucleotide sequences encoding them. The polynucleotide sequences given
 CC in AAA80606-AB0623 encode the 12 secreted protein sequences given in
 CC AAB5576-8B5593. The human secreted proteins have various activities
 CC dependent on the tissues in which they are expressed. Examples of the
 CC activities of the proteins include: immunosuppressant;
 CC anti-inflammatory; antiarthritic; antirheumatic; dermatological;
 CC antiproliferative; antiarteriosclerotic; anticancer; vulnary;
 CC antiviral; antibacterial; and antifungal activity. The proteins,
 CC polypeptides, agonists and antagonists may be used to treat prevent
 CC and/or diagnose various disease, disorders and conditions examples of
 CC which include: immune disorders e.g. Addison's disease, rheumatoid
 CC arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
 CC e.g. inflammatory bowel disease, Crohn's disease and nephritis;
 CC hyperproliferative disorders such as paraproelasma and purpura;
 CC cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
 CC cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
 CC sequences may also be used in wound healing and the treatment of
 CC infectious diseases. The human secreted protein gene #7 and protein
 CC protein gene #7 are represented in sequences AAA80612 and AAB5582. Secreted
 CC protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
 CC AAA80652-AB0661 represent genes which are related to the secreted protein
 CC gene#7.
 XX
 SQ Sequence 1034 AA;
 QY
 Query Match 1.1%; Score 13; DB 21; Length 1034;
 Best Local Similarity 100.0%; Pred. No. 0.0035;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 169 VIOLGNSNSIYPW 181
 166 VIOLGNSNSIYPW 178
 RESULT 32
 ABB90759
 ID ABB90759 standard; Protein; 1179 AA.
 XX
 AC ABB90759;
 XX
 DT 30-MAY-2002 (first entry)
 XX

DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 250.
 XX
 XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostratic;
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;
 KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;
 KW psoriasis.
 OS Homo sapiens.
 PN WO200210217-A2.
 PD 07-FEB-2002.
 PF 01-AUG-2001; 2001WO-US24031.
 PR 02-AUG-2000; 2000US-222599P.
 PR 11-AUG-2000; 2000US-224360P.
 PR 11-APR-2001; 2001US-282850P.
 XX (UYJO) UNIV JOHNS HOPKINS.
 PA
 PI St Croix B, Kinzler KW, Vogelstein B;
 DR WPI; 2002-291856/33.
 DR N-PsDB; ABL92113.
 XX
 PT An isolated molecule comprising an antibody variable region which
 PT specifically binds to an extracellular domain of a tumor endothelial
 PT marker (TEM) protein, useful for inhibiting tumor growth -
 PS Claim 35; Page 242-245; 331pp; English.
 XX
 CC The invention relates to an isolated molecule comprising an antibody
 CC variable region which specifically binds to an extracellular domain of a
 CC tumor endothelial marker (TEM) protein selected from ABB90732, ABB90740,
 CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
 CC proteins have cytostratic, immunostimulant and antiangiogenic activity.
 CC They are useful for inhibiting tumor growth, neoangiogenesis in
 CC subjects bearing a vascularised tumour, polycystic kidney disease,
 CC diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse
 CC and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and
 CC ABB90721-ABB90789) are disclosed, as are marker oligonucleotide
 CC sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and
 CC ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074;
 CC and pan-endothelial markers (PEM) ABL91903-ABL91995.
 XX
 SQ Sequence 1179 AA;
 QY
 Query Match 1.1%; Score 13; DB 23; Length 1179;
 Best Local Similarity 100.0%; Pred. No. 0.0039;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 169 VIOLGNSNSIYPW 181
 174 VIOLGNSNSIYPW 186
 RESULT 33
 ABB90788
 ID ABB90788 standard; Protein; 1180 AA.
 XX
 AC ABB90788;
 XX
 DT 30-MAY-2002 (first entry)
 DE Rat Tumour Endothelial Marker polypeptide SEQ ID NO 307.
 XX
 XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostratic;
 KW normal endothelial marker; pan-endothelial marker; immunostimulant;
 KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;
 KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;
 KW psoriasis.

```

XX OS Ratus sp..
XX XX WO200210217-A2.
XX PN
XX PD 07-FEB-2002.
XX XX
XX PF 01-AUG-2001, 2001WO-US24031.
XX XX
XX PR 02-AUG-2000, 2000US-222599P.
XX PR 11-AUG-2000, 2000US-224360P.
XX PR 11-APR-2001, 2001US-282850P.
XX XX
XX PA (UYUO ) UNIV JOHNS HOPKINS.
XX PT
XX PI St Croix B, Kinzler KM, Vogelstein B.
XX XX
XX DR WPI; 2002-291856/33.
XX DR N-PSDB; ABL92141.
XX XX
XX PT An isolated molecule comprising an antibody variable region which
XX PT specifically binds to an extracellular domain of a tumor endothelial
XX PT marker (TEM) protein, useful for inhibiting tumor growth -
XX PS Disclosure; Page 314-318; 331pp; English.
XX XX
XX CC The invention relates to an isolated molecule comprising an antibody
XX CC variable region which specifically binds to an extracellular domain of a
XX CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,
XX CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
XX CC proteins have cytostatic, immunostimulant and antiangiogenic activity.
XX CC They are useful for inhibiting tumour growth, neovascularogenesis in
XX CC subjects bearing a vascularised tumour, polycystic kidney disease,
XX CC diabetic retinopathy, rheumatoid arthritis and psoriasis. Human, mouse
XX CC and rat TEM genes and the encoded proteins (ABL92075-ABL92141 and
XX CC ABB90721-ABB90789) are disclosed, as are marker oligonucleotide
XX CC sequences: tumour endothelial markers (TEM) ABL91996-ABL92041 and
XX CC ABL92143-ABL92191; normal endothelial markers (NEM) ABL92042-ABL92074;
XX CC and pan-endothelial markers (PEM) ABL91903-ABL91995.
XX SQ
XX Sequence 1180 AA;
XX
XX Query Match 1.1%; Score 13; DB 23; Length 1180;
XX Best Local Similarity 100.0%; Pred. No. 0.0039;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 169 VIVLDGNSIYPM 181
XX DB 174 VIVLDGNSIYPM 186
XX
XX RESULT 34
XX AAU14231
XX ID AAU14231 standard; Protein; 1188 AA.
XX XX
XX AC AAU14231;
XX XX
XX DT 24-OCT-2001 (first entry)
XX XX
XX DE Human novel protein #102.
XX XX
XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
XX KW immunomodulatory; cytosatic; neuroprotective; vulnary; nootropic;
XX KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX KW tissue regeneration; immune disorder.
XX OS Homo sapiens.
XX XX
XX PN WO200155437-A2.

```

```

PD PD 02-AUG-2001.
XX PF
XX XX 25-JAN-2001; 2001WO-US02623.
XX XX
XX PR 25-JAN-2000; 2000US-0491404.
XX XX
XX PA (HYSE-) HYSEQ INC.
XX XX
XX PI Tang YT, Liu C, Dermanac RT;
XX XX
XX DR WPI; 2001-451939/48.
XX DR N-PSDB; AAS22536.
XX XX
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,
XX PT nervous system disorders, and for regenerating bone and cartilage -
XX PS Example 4; Page 578-581; 894pp; English.
XX XX
XX CC The invention relates to polynucleotides encoding novel human
XX CC proteins or their active domains. The polypeptides, polynucleotides and
XX CC antibodies raised against the polypeptides are used in a method of
XX CC treatment of a mammal and prevention of disorders caused by the aberrant
XX CC protein expression or activity. The polypeptides can be used as
XX CC molecular weight markers, food supplements, and in antibody production.
XX CC The polypeptides are used to identify compounds which bind to the
XX CC polypeptides. Polynucleotides of the invention are used as probes and
XX CC primers, for sequencing, for chromosome or gene mapping, in the
XX CC production of recombinant proteins, and in generating anti-sense DNA or
XX CC RNA and in gene therapy. Polypeptides of the invention can be used to
XX CC target drugs to a tumour, in assays to determine biological activity, to
XX CC raise antibodies/elicite an immune response, to determine quantitative
XX CC protein levels, as tissue markers, and to isolate receptors or ligands.
XX CC Polypeptides of the invention may also be useful in treating platelet
XX CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX CC ligament and/or nerve tissue, wound healing, treating burns, promoting
XX CC the proliferation, differentiation and survival of stem cells, as a
XX CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX CC fungal infection or from autoimmunity, cancer, allergy, asthma,
XX CC graft-versus-host disease, eczema, haemophilia, thrombosis,
XX CC anti-inflammatory diseases, nervous system disorders, and infection.
XX CC The present sequence represents a protein of the invention.
XX SQ
XX Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 169 VIVLDGNSIYPM 181
XX DB 166 VIVLDGNSIYPM 178
XX
XX RESULT 35.
XX AAU14467
XX ID AAU14467 standard; Protein; 1188 AA.
XX XX
XX AC AAU14467;
XX XX
XX DT 24-OCT-2001 (first entry)
XX XX
XX DE Human novel protein #338.
XX XX
XX KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
XX KW immunomodulatory; cytosatic; neuroprotective; vulnary; nootropic;
XX KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
XX KW tissue regeneration; immune disorder.
XX XX

```



```

OS Homo sapiens.
XX WO200155437-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001WO-US02623.
XX
XX 25-JAN-2000; 2000US-0491404.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-451939/48.
XX
XX N-PSDB; AAS22772.
XX
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
XX
XX Example 4; Page 828-831; 894pp; English.
XX
XX The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/ elicit an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence represents a protein of the invention.
XX
XX Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 169 VIVLDGNSNIYPW 181
DB 166 VIVLDGNSNIYPW 178

```

RESULT 36
AAB30929
ID AAB30929 standard; Protein; 1188 AA.

AC AAB30929;

DT 02-APR-2001 (first entry)

XX Amino acid sequence of a human alpha1 integrin chain.

XX Human; integrin; alpha1 subunit; fibroblast; muscle cell; chondrocyte;
XX osteoblast; stem cell; cell damage; muscle dystrophy; fibrosis;
XX wound healing; trauma; rheumatoid arthritis; osteoarthritis;
XX osteoporosis; cartilage damage; bone damage; cartilage.

```

XX OS Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..22
XX FT /note= "signal peptide"
XX FT 951..972
XX FT /note= "leucine zipper"
XX FT 1142..1164
XX FT /note= "transmembrane domain"
XX
XX WO200075187-A1.
XX
XX 14-DEC-2000.
XX
XX 31-MAY-2000; 2000WO-SE01135.
XX
XX 03-JUN-1999; 99SE-0002056.
XX
XX (ACT1-) ACTIVE BIOTECH AB.
XX
XX Gullberg D;
XX
XX WPI; 2001-071061/08.
XX
XX N-PSDB; AAC86871.
XX
XX Integrin subunit alpha 11 or integrin heterodimer comprising subunit
XX alpha 11 in association with subunit beta, useful for treating muscle
XX dystrophy, fibrosis, trauma, rheumatoid arthritis, and osteoarthritis -
XX
XX Disclosure; Fig 2a-c; 79pp; English.
XX
XX The present sequence represents a human integrin subunit, designated
XX alpha11. The alpha11 polynucleotide and polypeptide are useful as
XX markers of cell target molecules, such as fibroblasts, muscle cells,
XX chondrocytes, osteoblasts, mesenchymally derived cells or stem cells.
XX They are also used for determining the differential-stage of cells
XX during differentiation, development in pathological conditions, in
XX tissue regeneration, in transplantation or in therapeutic and
XX physiological repair of tissues. The pathological conditions involving
XX subunit alpha11 are selected from damage of cells, muscle dystrophy,
XX fibrosis, wound healing, trauma, rheumatoid arthritis, osteoarthritis
XX and osteoporosis, damage of cartilage and bone, and cartilage and bone
XX diseases. The polypeptide is useful for detecting the formation of
XX cartilage during embryonic development, for detecting physiological
XX therapeutic repair of cartilage and muscle, for selection and analysis,
XX or for sorting, isolating or purification of chondrocytes and muscle
XX cells, for detecting regeneration of cartilage or chondrocytes during
XX transplantation of cartilage or chondrocytes during transplantation of
XX cartilage or chondrocytes, respectively, or of muscle or muscle cells
XX during transplantation of muscle or muscle cells, respectively, and for
XX studies of differentiation or chondrocytes or muscle cells.
XX
XX Sequence 1188 AA;
XX
XX Query Match 1.1%; Score 13; DB 22; Length 1188;
XX Best Local Similarity 100.0%; Pred. No. 0.004;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
OY 169 VIVLDGNSNIYPW 181
DB 166 VIVLDGNSNIYPW 178

```

RESULT 37

ID AAB50085 standard; Protein; 1188 AA.

AC AAB50085;

DT 19-MAR-2001 (first entry)

XX Human A259.

```

XX Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KM kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
XX rheumatoid arthritis.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= Signal_peptide
FT 23..1188
FT Protein /label= Mature_protein
FT 1..1141
FT Domain /label= Extracellular_domain
FT 39..74
FT /label= Integrin_alphasubunit_repeat_domain_#1
FT 115..157
FT /label= Integrin_alphasubunit_repeat_domain_#2
FT 164..345
FT /label= I_domain
FT 367..392
FT Domain /label= Integrin_alphasubunit_repeat_domain_#3
FT 421..455
FT /label= Integrin_alphasubunit_repeat_domain_#4
FT 478..516
FT /label= Integrin_alphasubunit_repeat_domain_#5
FT 540..575
FT /label= Integrin_alphasubunit_repeat_domain_#6
FT 602..640
FT /label= Integrin_alphasubunit_repeat_domain_#7
FT 1142..1164
FT /label= Transmembrane_domain
FT 1165..1188
FT Domain /label= Cytoplasmic_domain
XX
XX WO200073339-A1.
XX
XX 07-DEC-2000.
XX
XX 15-MAY-2000; 2000WO-US13262.
XX
XX 28-MAY-1999; 99US-0322790.
XX 27-APR-2000; 2000US-0561263.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lora JM;
XX
XX WPI; 2001-041142/05.
XX N-PSDB; AAC91901, AAC91902.
XX
XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
XX diagnosis of fibrosis, e.g. of the liver
XX
XX Claim 8; Fig 1; 164pp; English.
XX
XX The present sequence is human integrin alpha subunit, A259. A259 is
XX homologous with the alpha1 and alpha10 integrin subunits and is
XX overexpressed in fibrosis. A259 is implicated in regulation of
XX proliferation, differentiation and/or function of many different cell
XX types. Inhibitors of A259 activity are useful for the treatment of liver
XX disease, particularly fibrosis, and also fibrosis in other organs
XX (specifically lung and kidney). In addition, A259 can be used for
XX treatment and prevention of cancer, osteoporosis, acute myeloid
XX leukaemia, HIV infection, and rheumatoid arthritis.
XX
XX Sequence 1188 AA;
SQ

```

```

DB 166 VIVLDGNSIYPW 178
|||||
RESULT 38
AAB50087
ID AAB50087 standard; Protein; 1188 AA.
XX
XX AAB50087;
AC
XX 19-MAR-2001 (first entry)
DT
XX
XX Murine A259.
XX
XX Human; A259; integrin alpha subunit; fibrosis; liver disease; lung;
KM kidney; cancer; osteoporosis; acute myeloid leukaemia; HIV infection;
XX rheumatoid arthritis.
XX
XX Mus sp.
XX
XX Location/Qualifiers
FH Key 1..22
FT Peptide /label= Signal_peptide
FT 23..1188
FT Protein /label= Mature_protein
FT 1..1141
FT Domain /label= Extracellular_domain
FT 39..74
FT /label= Integrin_alphasubunit_repeat_domain_#1
FT 115..157
FT /label= Integrin_alphasubunit_repeat_domain_#2
FT 164..345
FT /label= I_domain
FT 367..392
FT Domain /label= Integrin_alphasubunit_repeat_domain_#3
FT 421..455
FT /label= Integrin_alphasubunit_repeat_domain_#4
FT 478..516
FT /label= Integrin_alphasubunit_repeat_domain_#5
FT 540..575
FT /label= Integrin_alphasubunit_repeat_domain_#6
FT 602..640
FT /label= Integrin_alphasubunit_repeat_domain_#7
FT 1142..1164
FT /label= Transmembrane_domain
FT 1165..1188
FT Domain /label= Cytoplasmic_domain
XX
XX WO200073339-A1.
XX
XX 07-DEC-2000.
XX
XX 15-MAY-2000; 2000WO-US13262.
XX
XX 28-MAY-1999; 99US-0322790.
XX 27-APR-2000; 2000US-0561263.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Pan Y, Lora JM;
XX
XX WPI; 2001-041142/05.
XX N-PSDB; AAC91904, AAC91905.
XX
XX Nucleic acid encoding alpha-integrin subunits, useful for treatment and
XX diagnosis of fibrosis, e.g. of the liver
XX
XX Claim 8; Fig 5; 164pp; English.
XX
XX The present sequence is murine integrin alpha subunit, A259. A259 is
XX homologous with the alpha1 and alpha10 integrin subunits and is
XX overexpressed in fibrosis. A259 is implicated in regulation of
XX proliferation, differentiation and/or function of many different cell
XX

```

CC types. Inhibitors of A259 activity are useful for the treatment of liver
 CC disease, particularly fibrosis, and also fibrosis in other organs
 CC (specifically lung and kidney). In addition, A259 can be used for
 CC treatment and prevention of cancer, osteoporosis, acute myeloid
 CC leukaemia, HIV infection, and rheumatoid arthritis.

XX SQ Sequence 1188 AA;

Query Match 1.1%; Score 13; DB 22; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 VIVLDGNSNIYPM 181
 |||||
 166 VIVLDGNSNIYPM 178

RESULT 39

ID AAU10551 standard; Protein; 1188 AA.

XX AC AAU10551;

XX DI 14-FEB-2002 (first entry)

XX DE Human A259 polypeptide.

XX KW Human; A259: integrin alpha subunit; integrin alpha 10; secreted protein;
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;
 KW immune related disease; apoptotic disorder; neuronal tissue disease;
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
 KW antidiabetic; anticonvulsant; antiallergic; antiasthmatic; dermatological;
 KW antiparkinsonian.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..22

FT Domain /note= "Signal peptide"

FT Protein /note= "Extracellular domain"

FT Domain /note= "Mature human A259"

FT /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Integrin alpha repeat domain"

FT /note= "I domain or Von Willebrand Factor type A domain"

FT Domain /note= "Integrin alpha repeat domain"

FT /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Integrin alpha repeat domain. The specification
 states that this domain exists in human A549"

FT Domain /note= "Transmembrane domain"

FT /note= "Cyttoplasmic domain"

XX WO200181414-A2.

PD 01-NOV-2001.

XX PF 27-APR-2001; 2001WO-US13516.

XX PR 27-APR-2000; 2000US-0561263.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Pan Y, Lora J;

XX DR WPI; 2002-041397/05.

XX N-PSDB; AAS16873.

PT New A259 nucleic acids and polypeptides, which comprise integrin alpha
 subunit, useful for diagnosing, preventing or treating e.g. liver
 PT disease, kidney or lung fibrosis, cancers, blood disorders or immune
 PT related diseases

PS Claim 9; Fig 1; 168pp; English.

CC The invention relates to human and murine A259 nucleic acid molecules
 CC which encode secreted proteins with homology to integrin alpha subunits,
 CC specifically to integrin alpha 10. The A259 polypeptide and nucleic acid
 CC are useful for treating liver disease or fibrosis, particularly kidney
 CC fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also
 CC useful for diagnosing, preventing or treating cartilage and bone
 CC associated disorders (such as bone cancer, achondroplasia, myeloma,
 CC fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and
 CC osteoporosis), bone marrow, blood and haematopoietic disorders (such as
 CC acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune
 CC related diseases (such as HIV, viral infections, cancers, T cell
 CC autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.
 CC asthma and psoriasis), apoptotic disorders (such as systemic lupus
 CC erythematosus and insulin-dependent diabetes mellitus), diseases of the
 CC neuronal tissues (such as epilepsy and muscular dystrophy) and
 CC neurodegenerative diseases (such as Parkinson's disease and Huntington's
 CC disease). This sequence represents the human A259 polypeptide.

XX SQ Sequence 1188 AA;

Query Match 1.1%; Score 13; DB 23; Length 1188;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 VIVLDGNSNIYPM 181
 |||||
 166 VIVLDGNSNIYPM 178

RESULT 40

ID AAU10552 standard; Protein; 1188 AA.

XX AC AAU10552;

XX DI 14-FEB-2002 (first entry)

XX DE Murine A259 polypeptide.

XX KW Mouse; A259: integrin alpha subunit; integrin alpha 10; secreted protein;
 KW liver disease; fibrosis; lung; kidney; bone associated disorder; blood;
 KW cartilage associated disorder; haematopoietic disorder; bone marrow;
 KW immune related disease; apoptotic disorder; neuronal tissue disease;
 KW neurodegenerative disease; gene therapy; cancer; cytostatic; osteopathic;
 KW nephrotropic; immunomodulator; anti-inflammatory; neuroprotective;
 KW antidiabetic; anticonvulsant; antiallergic; antiasthmatic; dermatological;
 KW antiparkinsonian.

XX OS Mus musculus.

XX FH Key Location/Qualifiers

FT Peptide 1..22

FT /note= "Signal peptide"

FT	Domain	1..1141
FT	/note= "Extracellular domain"	
FT	Protein	23..1188
FT	/note= "Mature murine A259"	
FT	Domain	39..74
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	115..157
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	164..345
FT	/note= "I domain or Von Willebrand Factor type A domain"	
FT	Domain	367..392
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	421..455
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	478..516
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	540..575
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	602..640
FT	/note= "Integrin alpha repeat domain"	
FT	Domain	1142..1164
FT	/note= "Transmembrane domain"	
FT	Domain	1165..1188
FT	/note= "Cytoplasmic domain"	
PD	WO200181414-A2.	
PX	01-NOV--2001.	
PX	27-APR--2001; 2001WO-US13516.	
PX	27-APR--2000; 2000US-0561263.	
PR	(MILL-) MILLENNIUM PHARM INC.	
PA	Pan Y, Lora J;	
PI	WPI; 2002-041397/05.	
XX	N-PADB; AAS16874.	
XX	New A259 nucleic acids and polypeptides, which comprise integrin alpha	
PT	subunit, useful for diagnosing, preventing or treating e.g. liver	
PT	disease, kidney or lung fibrosis, cancers, blood disorders or immune	
PT	related diseases -	
PS	Claim 9; Fig 5; 1689P; English.	
CC	The invention relates to human and murine A259 nucleic acid molecules	
CC	which encode secreted proteins with homology to integrin alpha subunits,	
CC	specifically to integrin alpha 10. The A259 polypeptide and nucleic acid	
CC	are useful for treating liver disease or fibrosis, particularly kidney	
CC	fibrosis or lung fibrosis. The A259 polypeptide and nucleic acid are also	
CC	useful for diagnosing, preventing or treating cartilage and bone	
CC	associated disorders (such as bone cancer, achondroplasia, myeloma,	
CC	fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma and	
CC	osteoporosis), bone marrow, blood and haematopoietic disorders (such as	
CC	acute myeloid leukaemia, haemophilia, anaemia and thalassemia), immune	
CC	related diseases (such as HIV, viral infections, cancers, T cell	
CC	autoimmune disorders e.g. AIDS, and allergic inflammatory disorders e.g.	
CC	asthma and psoriasis), apoptotic disorders (such as systemic lupus	
CC	erythematosus and insulin-dependent diabetes mellitus), diseases of the	
CC	neuronal tissues (such as epilepsy and muscular dystrophy) and	
CC	neurodegenerative diseases (such as Parkinson's disease and Huntington's	
CC	disease). This sequence represents the murine A259 polypeptide.	
SQ	Sequence 1188 AA;	
Query Match	1.1%; Score 13; DB 23; Length 1188;	
Best Local Similarity	100.0%; Pred. NO. 0.004;	
Matches 13; Conservative	0; Mismatches 0; Indels 0; Gaps 0	
oy	169 VIVLDGNSNIYWP 181	

Db 166 V1VLDSNST1PEW 178

RESULT 41
ID AAB25582
AC AAB25582 standard; Protein; 1189 AA.
XX
XX AAB25582;
XX
XX 21-NOV-2000 (first entry)
XX
XX ITGAL1 protein encoded by human secreted protein gene #7.
XX
XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;
XX antirheumatic; dermatological; antiproliferative; antiarteriosclerotic;
XX anticancer; vulnery; antiviral; antibacterial; antifungal;
XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;
XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;
XX Crohn's disease; nephritis; hyperproliferative disorder;
XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;
XX melanoma; lymphoma; wound healing; human; chromosome 15 q22.3-23.
XX
XX Homo sapiens.
XX
XX WO200029435-A1.
XX
XX 25-MAY-2000.
XX
XX 27-OCT-1999; 99WO-US25031.
XX
XX 28-OCT-1998; 98US-0105971.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y,
XX Greene JW;
XX
XX WPI: 2000-387742/33.
XX
XX N-PSDB; AAB0612.
XX
XX
XX Isolated nucleic acid molecules encoding human secreted proteins are
XX used for the prevention, amelioration and treatment of autoimmune,
XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,
XX wounds, and infectious diseases -
XX
XX Claim 1; Figure 19A-F; 803pp; English.
XX
XX The present invention relates to 12 secreted human proteins and the
XX nucleotide sequences encoding them. The polynucleotide sequences given
XX in AAB0606-A80623 encode the 12 secreted protein sequences given in
XX AAB25576-B25593. The human secreted proteins have various activities
XX dependent on the tissues in which they are expressed. Examples of the
XX activities of the proteins include: immunosuppressant;
XX anti-inflammatory; antiarthritic; antirheumatic, dermatological;
XX antiproliferative; antiarteriosclerotic; anticancer; vulnery;
XX antiviral; antibacterial; and antifungal activity. The proteins,
XX polypeptides, agonists and antagonists may be used to treat prevent
XX and/or diagnose various disease, disorders and conditions examples of
XX which include: immune disorders e.g. Addison's disease, rheumatoid
XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders
XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;
XX hyperproliferative disorders such as paraproteinemias and purpura;
XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;
XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide
XX sequences may also be used in wound healing and the treatment of
XX infectious diseases. The human secreted protein gene #7 and protein
XX sequences are represented in sequences AAB0612 and AAB25582. Secreted
XX protein gene #7 is located at position chromosome 15 q22.3-23. Sequences
XX AAB0652-A80661 represent genes which are related to the secreted protein
XX gene#7.
XX
XX Sequence 1189 AA;
XX

Query Match 1.1%; Score 13; DB 21; Length 1189;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
 |||||
 Db 166 VIVLDGNSIYPM 178

RESULT 42

ABG12949
 ID ABG12949 standard; Protein; 1189 AA.

AC ABG12949;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #12940.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dymnac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS77136.

PT New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity

PS Claim 20; SEQ ID No 4308; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 1189 AA;

Query Match 1.1%; Score 13; DB 22; Length 1189;
 Best Local Similarity 100.0%; Pred. No. 0.004;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 181
 |||||
 Db 166 VIVLDGNSIYPM 178

RESULT 43

AAV07728
 ID AAV07728 standard; protein; 1183 AA.

AC AAV07728;

DT 01-JUL-1999 (first entry)

DE Armenian hamster alpha-1 integrin subunit protein.

KW VEGF; tumour angiogenesis inhibition; vascular endothelial growth factor;
 integrin cell surface receptor; capillary; blood vessel; hamster;
 alpha-1 subunit; alpha-2 subunit.

OS Cricetus migratorius.

PN WO9916465-A1.

PD 08-APR-1999.

PF 30-SEP-1997; 97WO-US17485.

PR 30-SEP-1997; 97WO-US17485.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (CLAF/) CLAFREY K P.

PA (DETM/) DETMAR M.

PA (SENG/) SENGER D R.

PI Clafey KP, Detmar M, Senger DR;

DR WPI; 1999-254930/21.

PT Inhibition of tumor angiogenesis through interaction of vascular
 endothelial growth factor and integrin cell surface receptors

PS Disclosure; Fig 2A-C; 64pp; English.

XX This invention describes a novel method for the inhibition of tumour
 CC angiogenesis mediated by vascular endothelial growth factor (VEGF) and
 CC integrin cell surface receptors expressed in vasculature of living
 CC subjects. The method inhibits new capillary and new blood vessel
 CC formation both within a tumour mass itself as well as in the immediately
 CC adjacent blood vasculature surrounding the perimeter of the tumour mass.
 CC Interaction and dependence upon VEGF to induce specific integrin
 CC heterodimers in tumour angiogenesis provides a novel method for
 CC inhibiting tumour angiogenesis, and unlike prior art relies on the
 CC specific inter-relationship of VEGF and integrins, rather than
 CC concentrating solely on one specific class of protein.

SO Sequence 1183 AA;

Query Match 1.0%; Score 12; DB 20; Length 1183;
 Best Local Similarity 100.0%; Pred. No. 0.037;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 VIVLDGNSIYPM 180
 |||||

Db 176 VIVLDGNSIYPM 187

RESULT 44

ABG12950
 ID ABG12950 standard; Protein; 117 AA.

AC ABG12950;

```

XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #12941.
DE
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM Food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX
XX MO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT.
PI
XX
XX WPI; 2001-639362/73.
DR
XX
XX N-PSDB; AAS77137.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnosis, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 43309; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes.
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 117 AA:
SQ

```

```

Query Match 0.9%; Score 11; DB 22; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1131 LGGLLALLLV 1141
    |||||
DB 62 LGGLLALLLV 72

```

```

RESULT 45
AAU09125
ID AAU09125 standard; protein; 185 AA.
XX
XX AAU09125;
AC
XX 16-JAN-2002 (first entry)
DT

```

```

XX Human integrin alpha2 I domain.
DE
XX
XX Human: integrin alpha2; I domain; protein co-ordinate data; thrombolytic;
KM cardiant; cerebroprotective; hepatotropic; antiinflammatory; vitucide;
KM vasotropic; antiarteriosclerotic; thrombotic disorder.
KM myocardial infarction; stroke; acute thrombosis; angioplasty.
KM coronary bypass grafting; liver fibrosis; liver necrosis; hepatitis;
KM arterial occlusion; restenosis; atherogenesis; anti-platelet therapy.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 17
FT /label= OTHER
FT /note= "Other
FT specification"
XX
XX WO200173444-A2.
PN
XX
XX 04-OCT-2001.
PD
XX
XX 27-MAR-2001; 2001WO-GB01358.
PF
XX
XX 28-MAR-2000; 2000US-192180P.
PR
XX
XX (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.
PA
XX
XX Farndale RW, Emsley J, Knight CG, Barnes MJ, Liddington RC;
PI
XX
XX WPI; 2001-648466/74.
DR
XX
XX New methods for identifying inhibitors of an I-domain-containing
PT polypeptide, particularly integrin, comprises use of coordinates of
PT peptide/receptor crystal structure -
XX
XX Example: Page 47; 100bp; English.
XX
XX The invention relates to methods for identifying a potential inhibitor of
CC an I-domain comprising polypeptide, particularly integrin, comprising
CC using a three dimensional structure of the integrin alpha2 I-domain as
CC defined in the specification to design or select a potential inhibitor.
CC Identified inhibitors are used to treat a disorder or disease. Such
CC disorders include thrombotic disorders, myocardial infarction and stroke,
CC acute thrombosis associated with angioplasty and coronary bypass
CC grafting, and with liver fibrosis or thrombotic complication of liver
CC necrosis after hepatitis. Inhibition of platelet alpha2beta1 may be used
CC to treat longer term occlusion of arteries, restenosis and atherogenesis.
CC Collagen receptor antagonism may be used as anti-platelet therapy.
CC The present sequence is the integrin alpha2 I domain.
XX
XX Sequence 185 AA:
SQ

```

```

Query Match 0.9%; Score 11; DB 22; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.069;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 272 VVVTGESHHDG 282
    |||||
DB 109 VVVTGESHHDG 119

```

```

RESULT 46
ABB72288
ID ABB72288 standard; Protein; 545 AA.
XX
XX ABB72288;
AC
XX
XX 04-APR-2002 (first entry)
DT
XX
XX Murine protein isolated from skin cells SEQ ID NO: 500.
DE
XX
XX Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KM

```

KW	developmental defect; inflammatory disease; dermatological; vulnery;
KV	immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX	
OS	Mus sp.
XX	
PN	WO200190357-A1.
XX	
PD	29-NOV-2001.
XX	
PF	24-MAY-2001; 2001WO-NZ00099.
XX	
PR	24-MAY-2000; 2000US-206650P.
PR	25-JUL-2000; 2000US-221232P.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	Watson JD, Strachan L, Sleeman M, Onrust R, Murison JG, Kumble KD;
XX	
DR	WPI; 2002-122020/16.
XX	
PT	New polynucleotides and polypeptides encoded by the polynucleotides
PT	isolated from skin cells, useful for treating skin wounds, cancers,
PT	growth and developmental defects, inflammatory diseases, or for
PT	modulating immune responses -
XX	
PS	Claim 4; Page 305-306; 466pp; English.
XX	
CC	The present invention provides the protein and coding sequences of cDNAs
CC	isolated from human, murine and rat skin cell libraries. The sequences
CC	can be used in the development of therapeutic agents useful in the
CC	treatment of skin diseases, including skin wounds, cancer, growth
CC	defects, developmental defects and inflammatory diseases. The proteins
CC	have important roles in the induction of hair growth, cell proliferation
CC	and cell-cell interaction, in maintaining tissue integrity, in wound
CC	healing and in modulating immune responses. The present sequence is a
XX	polypeptide of the invention.
XX	
SQ	Sequence 545 AA;
XX	
Query Match	0.9%; Score 11; DB 23; Length 545;
Best Local Similarity	100.0%; Pred. No. 0.18;
Matches 11; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1131 LGGULLLALLV 1141
DB	507 LGGULLLALLV 517
XX	
RESULT 47	
ABB72300	
ID	ABB72300 standard; Protein; 688 AA.
XX	
AC	ABB72300;
XX	
DT	04-APR-2002 (first entry)
XX	
DE	Rat protein isolated from skin cells SEQ ID NO: 624.
XX	
KW	Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW	developmental defect; inflammatory disease; dermatological; vulnery;
XX	immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
OS	Rattus sp.
XX	
PN	WO200190357-A1.
XX	
PD	29-NOV-2001.
XX	
PF	24-MAY-2001; 2001WO-NZ00099.
XX	
PR	24-MAY-2000; 2000US-206650P.
PR	25-JUL-2000; 2000US-221232P.
XX	

PA	(GENE-) GENESIS RES & DEV CORP LTD.
XX	
PI	Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG, Kumble KD,
DR	WPI; 2002-122020/16.
DR	N-PSDB; ABL34985.
XX	
PT	New polynucleotides and polypeptides encoded by the polynucleotides
PT	isolated from skin cells, useful for treating skin wounds, cancers,
PT	growth and developmental defects, inflammatory diseases, or for
PT	modulating immune responses
XX	
PS	Claim 4; Page 368-390; 466pp; English.
XX	
CC	The present invention provides the protein and coding sequences of cDNAs
CC	isolated from human, murine and rat skin cell libraries. The sequences
CC	can be used in the development of therapeutic agents useful in the
CC	treatment of skin diseases, including skin wounds, cancer, growth
CC	defects, developmental defects and inflammatory diseases. The proteins
CC	have important roles in the induction of hair growth, cell proliferation
CC	and cell-cell interaction, in maintaining tissue integrity, in wound
CC	healing and in modulating immune responses. The present sequence is a
CC	polypeptide of the invention.
XX	
SQ	Sequence 688 AA;
XX	
Query Match	0.9%; Score 11; DB 23; Length 688;
Best Local Similarity	100.0%; Pred. No. 0.22;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1131 LGGILLALLV 1141
DB	650 LGGILLALLV 660
XX	
RESULT 48	
ABB72289	
ID	ABB72289 standard; Protein; 696 AA.
XX	
AC	ABB72289;
XX	
DT	04-APR-2002 (first entry)
DE	Rat protein isolated from skin cells SEQ ID NO: 501.
XX	
KW	Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW	developmental defect; inflammatory disease; dermatological; vulnary;
KW	immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX	
OS	Rattus sp.
XX	
PN	WO200190357-A1.
PD	29-NOV-2001.
XX	
PF	24-MAY-2001; 2001WO-NZ00099.
XX	
PR	24-MAY-2000; 2000US-206650P.
PR	25-JUL-2000; 2000US-221232P.
XX	
PA	(GENE-) GENESIS RES & DEV CORP LTD.
PI	Watson JD, Strachan L, Sleeman M, Orrust R, Murison JG, Kumble KD;
XX	
DR	WPI; 2002-122020/16.
XX	
PT	New polynucleotides and polypeptides encoded by the polynucleotides
PT	isolated from skin cells, useful for treating skin wounds, cancers,
PT	growth and developmental defects, inflammatory diseases, or for
PT	modulating immune responses
XX	
PS	Claim 4; Page 306-307; 466pp; English.
XX	

CC The present invention provides the protein and coding sequences of cDNAs
 CC isolated from human, murine and rat skin cell libraries. The sequences
 CC can be used in the development of therapeutic agents useful in the
 CC treatment of skin diseases, including skin wounds, cancer, growth
 CC defects, developmental defects and inflammatory diseases. The proteins
 CC have important roles in the induction of hair growth, cell proliferation
 CC and cell-cell interaction, in maintaining tissue integrity, in wound
 CC healing and in modulating immune responses. The present sequence is a
 CC polypeptide of the invention.

XX
 SQ Sequence 696 AA;

Query Match 0.9%; Score 11; DB 23; Length 696;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLALLV 1141
 |||||
 DB 650 LGGLLLALLV 660

RESULT 49

AAW98238
 ID AAW98238 standard; Protein; 697 AA.

XX AAW98238;

XX 31-MAR-1999 (first entry)

XX H. pylori GHP0 702 protein.

XX GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KM peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX A1-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX13957.

XX New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases

XX Claim 8; Page 192-195; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHP0 protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.

XX Sequence 697 AA;

Query Match 0.9%; Score 11; DB 19; Length 697;
 Best Local Similarity 100.0%; Pred. No. 0.22;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1123 LWWIGSVLGG 1133
 |||||
 DB 128 LWWIGSVLGG 138

RESULT 50

ABG25584
 ID ABG25584 standard; Protein; 823 AA.

XX ABG25584;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #25575.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS89771.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 20; SEQ ID No 55943; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human

XX Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 823 AA;

Query Match 0.9%; Score 11; DB 22; Length 823;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 191 VVVTGDESHDG 201

RESULT 51

ABG29239 ID ABG29239 standard; Protein; 979 AA.

AC ABG29239;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #29230.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS93426.

PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity

PS Claim 20; SEQ ID No 59598; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 979 AA;

Query Match Best Local Similarity 0.9%; Score 11; DB 22; Length 979;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGDESHDG 282

Db 77 VVVTGDESHDG 87

RESULT 52

AAW07729 ID AAW07729 standard; protein; 1183 AA.

AC AAW07729;

DT 01-JUL-1999 (first entry)

DE Armenian hamster alpha-2 integrin subunit protein.

KW VEGF; tumour angiogenesis inhibitor; vascular endothelial growth factor;
integrin cell surface receptor; capillary; blood vessel; hamster;
alpha-1 subunit; alpha-2 subunit.

OS Cricetus migratorius.

PN WO916465-A1.

PD 08-APR-1999.

PF 30-SEP-1997; 97WO-US17485.

PR 30-SEP-1997; 97WO-US17485.

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (CLAF/) CLAFFEY K P.

PA (DETM/) DETMAR M.

PA (SENG/) SENG D R.

PI Claffey KP, Detmar M, Senger DR;

DR WPI; 1999-254930/21.

PT Inhibition of tumor angiogenesis through interaction of vascular
endothelial growth factor and integrin cell surface receptors

PS Disclosure; Fig 3A-C; 64pp; English.

XX This invention describes a novel method for the inhibition of tumour
XX angiogenesis mediated by vascular endothelial growth factor (VEGF) and
XX integrin cell surface receptors expressed in vasculature of living
XX subjects. The method inhibits new capillary and new blood vessel
XX formation both within a tumour mass itself as well as in the immediately
XX adjacent blood vasculature surrounding the perimeter of the tumour mass.
XX Interaction and dependence upon VEGF to induce specific integrin
XX heterodimers in tumour angiogenesis provides a novel method for
XX inhibiting tumour angiogenesis, and unlike prior art relies on the
XX specific inter-relationship of VEGF and integrins, rather than
XX concentrating solely on one specific class of protein.

SO Sequence 1183 AA;

Query Match Best Local Similarity 0.9%; Score 11; DB 20; Length 1183;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 VVVTGDESHDG 282

Db 279 VVVTGDESHDG 289

RESULT 53

AAW0542 ID AAW0542 standard; Protein; 1367 AA.

AC AAW0542;

DT 26-JAN-1999 (first entry)

DE Integrin alpha-2 chain.

XX 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 PR (HYSE-) HYSEQ INC.
 PA Drmanac RT, Liu C, Tang YT;
 XX WPI, 2001-639362/73.
 DR N-PSDB; AAS70024.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnosis, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 36196; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1223 AA;
 OY
 Db 483 GSYFGSELCP 492
 552 GSYFGSELCP 561
 RESULT 56
 AAY32249
 ID AAY32249 standard; Peptide; 13 AA.
 AC AAY32249;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Bovine integrin subunit alpha-10 cryptic peptide 5.
 XX
 KM Integrin alpha-10; Isal0; human; trauma; rheumatoid arthritis;
 KM osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KM inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KM fibroblast; vaccine; marker.
 XX
 OS Bos taurus.
 XX
 PF Key Location/Qualifiers
 FH

FT Misc-difference 10
 FT /note= "unidentified"
 XX
 XX WO9951639-A1.
 PN 14-OCT-1999.
 XX
 PD 31-MAR-1999; 99WO-SE00544.
 XX
 PF 02-APR-1998; 98SE-0001164.
 XX
 PR 28-JAN-1999; 99SE-0000319.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 PI Lundgren-Akerlund E;
 XX
 DR WPI, 2000-052639/04.
 XX
 PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation -
 XX
 PS Example 2; Fig 2; 90pp; English.
 XX
 CC This sequence represents tryptic peptide 5 of integrin subunit
 CC alpha-10 (Isal10) isolated from bovine chondrocytes. 6 Tryptic
 CC peptides (see AAY32245-50) were obtained. PCR primers (see
 CC AAY324721-24) based on peptides 1 and 2 were used to obtain bovine
 CC Isal10 cDNA fragments. These were used to screen a human
 CC articular chondrocyte cDNA library in order to isolate human Isal10
 CC cDNA (see AAY34719). The invention relates to a recombinant or
 CC associated integrin heterodimer comprising the alpha-10 subunit in
 CC or the subunit alpha-10 can be used as a marker or target of all
 CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
 CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC CC
 CC compositions and vaccines.
 CC
 SQ Sequence 13 AA;
 OY
 Db 700 FTASLDEWT 708
 1 FTASLDEWT 9
 RESULT 57
 AAM35852
 ID AAM35852 standard; protein; 1065 AA.
 AC AAM35852;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Human CD11 for use in T lymphocyte veto molecule.
 XX
 KM Human; CD11; T lymphocyte veto molecule; chimeric molecule;
 KM targeting polypeptide; suppression; immune response; treatment;
 KM autoimmune disease; allergy; immunological disorder;
 KM transplant rejection.
 XX
 OS Homo sapiens.
 XX
 PN WO9737687-A1.
 XX
 PD 16-OCT-1997.
 XX
 PF 10-APR-1997; 97WO-US05943.
 XX

PR 10-APR-1996; 96US-0630172.
 XX (NAJE-) NAT JEWISH CENT IMMUNOLOGY & RESPIRATORY.
 PA
 XX
 PI Staez UD;
 XX
 DR WPI; 1997-512419/47.
 XX
 PT T lymphocyte veto molecule comprising response cell activating
 PT protein - linked to molecule that targets stimulator cell marker.
 PT used for selective suppression of immune response, e.g. prevention
 PT of graft rejection or treatment of auto-immune disease
 XX
 PS Claim 37; Pages 61-65; 309pp; English.
 XX
 CC A novel T lymphocyte veto molecule is a chimeric molecule
 CC comprising a protein, e.g. the present sequence, linked to a
 CC targeting polypeptide that binds a molecule, which differentiates
 CC a host cell from a tissue graft cell, or selectively targets a
 CC stimulator cell involved in the autoimmune response.
 CC A veto molecule, in which the protein binds a molecule that targets
 CC stimulator cells, can be used to suppress an immune response and
 CC therefore treat autoimmune diseases, e.g. systemic lupus
 CC erythematosus, myasthenia gravis, rheumatoid arthritis, insulin
 CC dependent diabetes mellitus, multiple sclerosis, coeliac disease,
 CC autoimmune thyroiditis, Addison's or Grave's diseases and
 CC rheumatoid carditis, allergies and other immunological disorders.
 CC Where the protein binds a molecule that differentiates graft and
 CC host cells, the veto molecule can be used to reduce transplant
 CC rejection. The veto molecule provides specific regulation of
 CC particular stimulator cells that can kill graft cells or respond
 CC to autoantigens, but leave other stimulator cells unaffected, e.g.
 CC CD4 or CD8 positive cells can be regulated without one affecting
 CC the other. The veto molecule can be administered locally to
 CC minimise generalised immunosuppression.
 XX
 SQ Sequence 1065 AA;
 XX
 QY Query Match 0.8%; Score 9; DB 18; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 566 DVAVGAPLE 574
 |||||
 514 DVAVGAPLE 522

XX
 XX Nucleic acid encoding soluble LFA-1 protein - used to identify
 PT anti-inflammatory agents for use with LFA-1 mediated activity
 PT
 XX
 PS Disclosure; Page 13; 14pp; Japanese.
 XX
 CC This sequence represents a soluble LFA-1 alpha chain protein. The
 CC encoding nucleic acids can be used for the evaluation of an
 CC anti-inflammatory agent by a test for the combination between
 CC ICAM protein and a soluble LFA-1 hetero dimer protein. The soluble
 CC LFA-1 protein does not associate in the absence of a surfactant.
 XX
 SQ Sequence 1086 AA;
 XX
 QY Query Match 0.8%; Score 9; DB 20; Length 1086;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 566 DVAVGAPLE 574
 |||||
 538 DVAVGAPLE 546

RESULT 59
 AAB03973
 ID AAB03973 standard; Protein; 1145 AA.
 AC AAB03973;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Alpha_L chain I domain of leukocyte function associated antigen.
 XX
 KW Leukocyte function associated antigen; LFA-1; I domain; inhibitor;
 KW ligand; binding; pathology; treatment; screening; detection;
 KW inflammatory disease; autoimmune disease; reperfusion injury;
 KW myocardial infarction; stroke; organ transplant;
 KW respiratory distress syndrome; glomerulonephritis; Crohn's disease;
 KW T cell mediated disease; rheumatoid arthritis; osteoarthritis;
 KW spondylitis; thyroid associated ophthalmopathy; Behcet disease;
 KW sepsis; asthma; chronic bronchitis; silicosis; pulmonary sarcoidosis;
 KW fibrosis; cystic fibrosis; keloid formation; scar formation;
 KW atherosclerosis; transplant rejection; lupus;
 KW inflammatory bowel disease; ulcerative colitis; leukaemia;
 KW atopic dermatitis; psoriasis; urticaria; uveitis; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200060355-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 03-APR-2000; 2000WO-US08841.
 XX
 PR 02-APR-1999; 99US-0285477.
 XX
 PA (ICOS-) ICOS CORP.
 PA (ABBO) ABBOTT LAB.
 XX
 PI Staunton D, Van Der Vieren M, Hutch J, Fowler K, Orme M;
 PI Olejniczak ET;
 XX
 DR WPI; 2000-665028/64.
 N-PSDB; AAA54333.
 XX
 PT Identifying inhibitor of leukocyte function-associated antigen binding
 PT to natural ligand for ameliorating inflammatory diseases, by measuring
 PT antigen-ligand binding in presence and absence of test compound
 XX
 PS Example 2; Page 51-54; 66pp; English.
 XX
 CC Inhibitors of leukocyte function-associated antigen binding are
 CC useful for manufacturing a medicament for ameliorating a pathology

CC arising from LFA-1 binding to its ligand such as inflammatory
 CC diseases, autoimmune diseases, reperfusion injury, myocardial
 CC infarction, stroke, organ transplant, adult respiratory distress
 CC syndrome, acute glomerulonephritis, Crohn's disease, T cell mediated
 CC diseases, rheumatoid arthritis, osteoarthritis, spondylitis, thyroid
 CC associated ophthalmopathy, Behcet disease, sepsis, asthma, chronic
 CC bronchitis, allergic respiratory distress syndrome, chronic pulmonary
 CC inflammatory disease, such as chronic obstructive pulmonary disease,
 CC silicosis, pulmonary sarcoidosis, fibrosis, cystic fibrosis, keloid
 CC formation, scar formation, atherosclerosis, transplant rejection
 CC disorders, such as graft versus host reaction and allograft
 CC rejection, chronic glomerulonephritis, lupus, inflammatory bowel
 CC disease, such as ulcerative colitis, proliferative lymphocyte
 CC diseases, such as leukemia, and inflammatory dermatoses, such as
 CC atopic dermatitis, psoriasis, urticaria, and uveitis.

SO Sequence 1145 AA;

Query Match 0.8%; Score 9; DB 21; Length 1145;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 |||||
 Db 513 DVAVGAPLE 521

RESULT 60
 AAR05782
 ID AAR05782 standard; Protein; 1170 AA.
 XX AAR05782;
 XX
 DT 23-AUG-1990 (first entry)
 XX
 DE LFA-1 alpha subunit.
 XX
 KM Lymphocyte function associated antigen; inflammation; metastasis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT Domain /label= sig_peptide
 FT Domain 26..108
 FT Domain /label= extracellular_domain
 FT Domain 1089..1117
 FT Region /label= transmembrane_domain
 FT Peptide 1118..1170
 FT Peptide /label= cytoplasmic_tail
 FT Peptide 118..132
 FT Peptide /label= tryptic_fragment
 FT Peptide 226..237
 FT Peptide /label= tryptic_fragment
 FT Peptide 282..288
 FT Peptide /label= tryptic_fragment
 FT Peptide 433..441
 FT Peptide /label= tryptic_fragment
 FT Peptide 521..531
 FT Peptide /label= tryptic_fragment
 FT Peptide 559..582
 FT Peptide /label= tryptic_fragment
 FT Peptide 590..604
 FT Peptide /label= tryptic_fragment
 FT Peptide 831..844
 FT Peptide /label= tryptic_fragment
 FT Peptide 957..974
 FT Peptide /label= tryptic_fragment
 FT Modified-site 65..67
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 89..91
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 188..190
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 649..651

FT Modified-site /label= N-glycosylation_site
 FT Modified-site 670..672
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 726..728
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 730..732
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 862..864
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 885..887
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 897..899
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 1060..1062
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 1071..1073
 FT Modified-site /label= N-glycosylation_site
 FT Modified-site 1140
 FT Modified-site /label= serine_phosphorylation_site
 FT Modified-site 1145
 FT Modified-site /label= serine_phosphorylation_site
 FT Modified-site 1163
 FT Modified-site /label= serine_phosphorylation_site
 FT Modified-site 1165
 FT Modified-site /label= serine_phosphorylation_site

PN EP362526-A.
 PD 11-APR-1990.
 XX
 XX 17-AUG-1989; 89EP-0115160.
 XX
 PR 23-AUG-1988; 88US-0235227.
 PR 09-MAR-1989; 89US-0321017.
 XX
 PA (DANA-) DANA FARBER CANCER.
 PI Springer TA, Larson R;
 PI
 DR WPI; 1990-108985/15.
 DR N-PSDB; AAO03842.
 XX
 PT Pure alpha subunit of lymphocyte function associated antigen -
 PT and encoding DNA sequences, useful eg for suppressing
 PT inflammation or metastasis
 XX
 PS Disclosure; Fig 3; 27pp; English.
 XX
 CC The alpha-subunit (a-SU) can bind to ICAM-1 (or other natural
 CC ligands) on the surface of cells, and can associate with the beta-SU
 CC to form a heterodimer (also able to bind to ICAM-1). a-SU, and its
 CC derivs., are useful in suppressing inflammation, metastasis and its
 CC growth of a-SU expressing tumour cells and is used in the treatment
 CC of viral infections.
 CC The pref. dose is 1 pg - 10 mg/kg.
 XX

SO Sequence 1170 AA;

Query Match 0.8%; Score 9; DB 11; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 |||||
 Db 538 DVAVGAPLE 546

RESULT 61
 AAR80107
 ID AAR80107 standard; Protein; 1170 AA.
 XX AAR80107;
 XX

DT 04-MAR-1996 (first entry)
 XX LFA-1 alpha subunit CD11a.
 DE CD11a; leucocyte function-associated antigen-1; LFA-1; integrin.
 XX
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Peptide 1..25 /label= Sig_peptide
 FT Region 32..79 /label= Repeat
 FT FT /note= "Repeat I"
 FT Region 82..132 /label= Repeat
 FT FT /note= "Repeat II"
 FT Region 339..391 /label= Repeat
 FT FT /note= "Repeat III"
 FT Region 392..446 /label= Repeat
 FT FT /note= "Repeat IV"
 FT Region 447..508 /label= Repeat
 FT FT /note= "Repeat V"
 FT Region 509..567 /label= Repeat
 FT FT /note= "Repeat VI"
 FT Region 568..629 /label= Repeat
 FT FT /note= "Repeat VII"
 FT Domain 170..349 /label= IDomain
 FT FT /note= "internal domain"
 FT Domain 1089..1112 /label= Trans
 FT FT /note= "transmembrane domain"
 FT Domain 1113..1170 /label= Cyto
 FT FT /note= "cytoplasmic domain"
 XX
 PN WO9528170-A1.
 XX
 PD 26-OCT-1995.
 XX
 PF 19-APR-1995; 95WO-US04886.
 XX
 PR 19-APR-1994; 94US-0229513.
 XX
 PA (UNIV) UNIV KANSAS.
 XX
 PI Benedict S, Chan MA, Siahaan TJ, Tibbets SA;
 XX
 DR WPI; 1995-373631/48.
 XX
 PT Changing peptide reactivity via conjugation with a second peptide
 PT causes change in conformation of first peptide, pref. LFA-1 and
 PT ICM-1 functional domain derived peptide(s)
 XX
 PS Claim 11; Page 34-37; 64pp; English.
 XX
 CC Functional domains derived from LFA-1 alpha subunit CD11a (AAR80107),
 CC LFA-1 beta subunit CD8 (AAR80107) and ICM-1 CD43 (AAR80110) are used
 CC as the basis of short, synthetic peptides (AAR80109, AAR80111-24) that
 CC modulate ICM/LFA binding interaction. CD11a is obcd. from human
 CC PMA-stimulated T-lymphocyte HL-60 cells.
 XX
 SO Sequence 1170 AA;

QY 566 DVAVGAPLE 574
 DB 538 DVAVGAPLE 546
 RESULT 62
 ABB76377
 ID ABB76377 standard; Protein; 1170 AA.
 XX
 AC ABB76377;
 DT 22-AUG-2002 (first entry)
 XX
 DE Lymphocyte function associated antigen-1.
 XX
 KW Lymphocyte function associated antigen-1; LFA-1; human;
 KW alpha/beta protein; ligand binding.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Domain 153..332 /label= I_domain
 FT FT
 PN WO200231511-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 12-OCT-2001; 2001WO-US32047.
 XX
 PR 12-OCT-2000; 2000US-239750P.
 XX
 PA (ICOS-) ICOS CORP.
 XX
 PI Stauton DE;
 XX
 DR WPI; 2002-471361/50.
 XX
 PT Modulating binding interactions between alpha/beta proteins comprising
 PT an allosteric regulatory sites, and a binding partner, by contacting with
 PT an allosteric effector molecule that interacts with the regulatory site
 XX
 PS Disclosure; Fig 2; 163pp; English.
 XX
 CC The present sequence is the protein sequence of lymphocyte function
 CC associated antigen-1 (LFA-1), which includes an integrin (I)
 CC domain. The invention provides methods of modulating the binding
 CC interaction of a first molecule, which is not LFA-1 or an I
 CC domain-containing fragment of LFA-1, and a binding partner. The
 CC first molecule has an alpha/beta domain structure comprising an
 CC allosteric regulatory site. Modulation of the binding interaction
 CC involves contacting with an allosteric effector molecule that
 CC interacts with the regulatory site. In vivo methods are expected
 CC to alleviate and/or prevent pathological states arising from
 CC aberrant binding activity. Methods for identifying modulators are
 CC also provided.
 XX
 SO Sequence 1170 AA;

Query Match 0.8%; Score 9; DB 23; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 DB 538 DVAVGAPLE 546
 RESULT 63
 AA080251
 ID AA080251 standard; Protein; 1170 AA.

XX AC AAU80251;
 XX 15-JUL-2002 (first entry)
 XX
 DE Human integrin 1 alpha-L subunit protein.
 XX
 KW Integrin; antiinflammatory; immunosuppression; nephritis; dermatitis;
 KW inflammatory disease; autoimmune disorder; Crohn's disease;
 KW human immunodeficiency virus; HIV; myocardial infarction;
 KW Sjogren's syndrome; rheumatoid arthritis.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 214
 FT /note= "Encoded by TGC"
 PN WO200218583-A2.
 PD 07-MAR-2002.
 PF 31-AUG-2001; 2001WO-US27227.
 PR 01-SEP-2000; 2000US-229700P.
 XX
 PA (BLOO-) CENT BLOOD RES INC.
 XX
 PI Springer TA, Shimoaka M, Lu C;
 DR WPI; 2002-382964/41.
 DR N-PSDB; ABR50045.
 XX
 PT Modified integrin-I or integrin I-like domain polypeptide useful as an
 PT immunogen to produce antibodies specific to polypeptide, comprises a
 PT disulfide bond such that polypeptide is stabilized in a desired
 PT conformation
 XX
 PS Disclosure; Page 105-108; 112pp; English.
 XX
 CC This invention relates to a modified integrin-I or integrin I-like
 CC domain polypeptide comprising at least one disulfide bond so that the
 CC domain is stabilized in a desired conformation. The polypeptide of
 CC the invention may have antiinflammatory or immunosuppressive activities.
 CC The polypeptides of the invention have an open conformation and are
 CC useful as immunogens to produce antibodies that selectively bind to
 CC integrin I-domain, and for identifying a modulator of integrin activity,
 CC or of interaction of an integrin and a cognate ligand. The polypeptide
 CC of the invention, or antibodies (preferably anti-LFA-1 antibody)
 CC is useful for treating or preventing an integrin mediated disorder which
 CC is an inflammatory or autoimmune disorder in a subject and for
 CC inhibiting the binding of an integrin to a cognate ligand in a subject.
 CC A therapeutic composition comprising the peptide of the invention is
 CC useful for treating an integrin mediated disorder such as Crohn's
 CC disease, nephritis; human immunodeficiency virus (HIV), myocardial
 CC infarction, Sjogren's syndrome, rheumatoid arthritis, dermatitis. The
 CC polypeptides and/or active or antigenic fragments are useful as
 CC reagents for diagnosis of integrin-mediated disorders. The present
 CC sequence represents the human integrin-1 alpha-L protein subunit used to
 CC generate the mutant polypeptides of the invention.
 XX
 SQ Sequence 1170 AA;

Query Match 0.8%; Score 9; DB 23; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
 |||||
 DB 538 DVAVGAPLE 546

RESULT 64

AAV32248
 ID AAV32248 standard; peptide; 8 AA.
 XX
 AC AAV32248;
 XX
 DT 15-FEB-2000 (first entry)
 XX
 DE Bovine integrin subunit alpha-10 tryptic peptide 4.
 XX
 KW Integrin alpha-10; ISa10; human; trauma; rheumatoid arthritis;
 KW osteoarthritis; osteoarthritis; cancer; atherosclerosis;
 KW inflammation; therapy; cartilage; chondrocyte; osteoblast;
 KW fibroblast; vaccine; marker.
 XX
 OS Bos taurus.
 XX
 PN WO951639-A1.
 PD 14-OCT-1999.
 PF 31-MAR-1999; 99WO-SE00544.
 PR 02-APR-1998; 98SE-0001164.
 PR 28-JAN-1999; 99SE-0000319.
 XX
 PA (ACTI-) ACTIVE BIOTECH AB.
 XX
 PI Lundgren-Akerlund E;
 DR WPI; 2000-052639/04.
 XX
 PT New isolated integrin subunit alpha-10, used as a marker or target
 PT molecule for cells during development, regeneration and pathological
 PT conditions, e.g. arthritis, osteoarthritis, cancer, atherosclerosis or
 PT inflammation
 XX
 PS Example 2; Fig 2; 90pp; English.
 XX
 CC This sequence represents tryptic peptide 4 of integrin subunit
 CC alpha-10 (ISa10) isolated from bovine chondrocytes. 6 tryptic
 CC peptides (see AAV32245-50) were obtained. PCR primers (see
 CC AA234721-24) based on peptides 1 and 2 were used to obtain bovine
 CC ISa10 cDNA fragments. These were used to screen a human
 CC articular chondrocyte cDNA library in order to isolate human ISa10
 CC cDNA (see AA234719). The invention relates to a recombinant or
 CC isolated integrin heterodimer comprising the alpha-10 subunit in
 CC association with subunit beta (especially beta-1). The heterodimer
 CC or the subunit alpha-10 can be used as a marker or target of all
 CC types of cells, e.g. of chondrocytes, osteoblasts and fibroblasts.
 CC They can also be used as active ingredients in pharmaceutical
 CC compositions and vaccines.
 XX
 SQ Sequence 8 AA;

Query Match 0.7%; Score 8; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 551 FAMGALPD 558
 |||||
 DB 1 FAMGALPD 8

RESULT 65
 AAM99005
 ID AAM99005 standard; peptide; 17 AA.
 XX
 AC AAM99005;
 XX
 DT 11-MAY-1999 (first entry)
 XX
 DE I domain peptide P10.
 XX

KW Cyclic integrin binding peptide; integrin alpha-2i domain; inhibitor;
 KW collagen I; collagen IV; laminin-1; cell migration; cancer;
 KW cardiovascular disease; periodontal disease.
 XX
 OS Synthetic.
 XX
 PN WO902551-A1.
 XX
 PD 21-JAN-1999.
 XX
 PF 09-JUL-1998; 98WO-F100579.
 XX
 PR 11-JUL-1997; 97US-0893526.
 XX
 PA (HEIN/) HEINO J.
 PA (IVAS/) IVASKA J.
 PA (KAEP/) KAEPYLAIE J.
 XX
 PI Heino J, Ivaska J, Kaepylae J;
 XX
 DR WPI, 1999-120775/10.
 XX
 PT Cyclic integrin binding peptides - used to inhibit
 PT integrin-dependent cell migration
 XX
 PS Example 3; Page 25; 59pp; English.
 XX
 CC The present sequence invention describes cyclic integrin binding peptides
 CC comprising the amino acid sequence RKK, preferably RKKH. Also described
 CC in the present invention are: (1) a cyclic peptide comprising the amino
 CC acid sequence X1RKKHX2n where X is any amino acid and n=1-4; (2) a
 CC cyclic integrin binding peptide comprising the amino acid sequence
 CC CTRKKHDMC or CTRKKHDMQC; (3) a pharmaceutical composition comprising
 CC one of the above integrin binding peptides; and (4) a binding assay for
 CC identifying integrin binding agents, comprising: (a) biotinylating the
 CC integrin binding agent to be assayed; (b) reacting the biotinylated
 CC agent with an immobilised recombinant alpha 2i domain or domain-derived
 CC peptides in conditions suitable for binding; (c) washing the solid
 CC support carrying the bound agent; (d) adding a labeled biotin-binding
 CC agent; and (e) detecting any bound integrin binding agent. The integrin
 CC binding peptides can be used for inhibiting integrins. They can also be
 CC used for inhibiting integrin-dependent cell migration, such as
 CC associated with cancer, cardiovascular disease or a periodontitis
 CC condition. They can also be used for inhibiting the adhesion of
 CC platelets to collagen or collagen induced platelet aggregation in a
 CC patient. The present sequence represents a peptide from the present
 CC invention.
 XX
 SQ Sequence 17 AA;
 XX
 QY Query Match 0.7%; Score 8; DB 20; Length 17;
 Db Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 275 TDGESHG 282
 Db 1 TDGESHG 8
 XX
 RESULT 66
 AAY12862
 ID AAY12862 standard; Protein; 47 AA.
 XX
 AC AAY12862;
 XX
 DT 21-JUN-1999 (first entry)
 XX
 DE Human 5' EST secreted protein SEQ ID NO:452.
 XX
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW

KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 XX
 OS Homo sapiens.
 XX
 PN WO906549-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-1B01231.
 XX
 PR 01-AUG-1997; 97US-0905279.
 XX
 PA (GEST) GENSET.
 XX
 PI Duclert A, Dumae Milne Edwards J, Lacroix B;
 XX
 DR WPI, 1999-153779/13.
 DR N-PSDB; AAX51640.
 XX
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT CDNA libraries derived from testis, ovary, uterus and spleen tissue
 XX
 PS Claim 34; Page 490; 522pp; English.
 XX
 CC AAX51459 to AAX51691 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY12681 to
 CC AAY12913, respectively. The proteins given represent the signal peptide
 CC and an N-terminal fragment of a secreted protein. The nucleic acid
 CC sequences can be used for producing secreted human gene products. They
 CC can also be used to develop products for diagnosis and therapy. The
 CC proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.
 XX
 SQ Sequence 47 AA;
 XX
 QY Query Match 0.7%; Score 8; DB 20; Length 47;
 Db Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1133 GULLLALL 1140
 Db 12 GULLLALL 19
 XX
 RESULT 67
 AAG00326
 ID AAG00326 standard; Protein; 66 AA.
 XX
 AC AAG00326;
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 4407.
 XX
 KW Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 PD 06-SEP-2000.
 XX

PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 98US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR N-PSDB; AAC00332.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
PS Claim 13; SEQ ID 4407; 71bp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 66 AA;
XX
Query Match 0.7%; Score 8; DB 21; Length 66;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1133 GULLLALL 1140
DB 12 GULLLALL 19
XX
RESULT 68
AAAY41696
ID AAAY41696 standard; Protein; 67 AA.
XX
AC AAAY41696;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO617 protein sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridization;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.
XX
OS Homo sapiens.
XX
PN WO9346281-A2.
XX
PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
XX 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077641.
PR 12-MAR-1998; 98US-0077649.
PR 13-MAR-1998; 98US-0077791.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.

PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0081070.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081836.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082566.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 23-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0083366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.

```

PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
XX
XX (GETH ) GENENTECH INC.
XX
XX
XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
DR WPI; 1999-551358/46.
DR N-PSDB; AAZ33963.
XX
XX New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX Claim 12; Fig 33; 530pp; English.
XX
XX The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ3891 to
CC AAZ4338, and AAZ4165 to AAZ4174 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
XX Sequence 67 AA;
SQ
XX
XX Query Match 0.7%; Score 8; DB 20; Length 67;
XX Best Local Similarity 100.0%; Pred.No. 24;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1133 GLLLIALL 1140
XX |||||||
XX 12 GLLLIALL 19
DB
XX
XX RESULT 69
XX AAB44252
XX ID AAB44252 standard; Protein; 67 AA.
XX
XX AAB44252;
XX AC
XX XX
XX 08-FEB-2001 (first entry)
XX
XX DE Human PRO617 (UNQ353) protein sequence SEQ ID NO:85.
XX
XX Human; secreted protein; transmembrane protein; PRO; EST; cytoslatic;
XX expressed sequence tag; detection; cancer.
XX
XX Homo sapiens.
XX OS
XX PN WO200053756-A2.
XX
XX 14-SEP-2000.
XX
XX 18-FEB-2000; 2000WO-US04341.
XX
XX 08-MAR-1999; 98WO-US05028.
XX 12-MAR-1999; 98US-0123957.
XX 29-MAR-1999; 98US-0126773.
XX 21-APR-1999; 98US-0130232.
XX 28-APR-1999; 98US-0131445.
XX 14-MAY-1999; 98US-0134287.
XX 23-JUN-1999; 98US-0141037.
XX 26-JUL-1999; 98US-014698.
XX 29-OCT-1999; 98US-0162506.
XX

```

PR	30-NOV-1999;	99MO-US28313.
PR	02-DEC-1999;	99MO-US28551.
PR	02-DEC-1999;	99MO-US28565.
PR	16-DEC-1999;	99MO-US30095.
PR	30-DEC-1999;	99MO-US31243.
PR	30-DEC-1999;	99MO-US31274.
PR	05-JAN-2000;	2000MO-US00219.
PR	06-JAN-2000;	2000MO-US00277.
PR	06-JAN-2000;	2000MO-US00376.
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL,	
PI	Ferrare N, Filvarsoff E, Fong S, Gao W, Gerber H, Gerritsen ME;	
PI	Goddard A, Godowski PJ, Grimaldi CJ, Gunney AL, Hillan KJ;	
PI	Klajavins IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;	
PI	Shelton DL, Stewart TA, Tunas D, Williams PM, Wood WT;	
XX		
DR	WPI: 2000-611443/58.	
DR	N-PSDB; AAC78478.	
XX		
PT	Novel PRO polypeptides and polynucleotides used in detection methods,	
PT	to target bioactive molecules to specific cells, and to modulate	
PT	cellular activities -	
XX		
PS	Claim 12; Fig 33; 636pp; English.	
XX		
CC	AAC78458 to AAC78599 represent polynucleotide and EST (expressed	
CC	sequence tag) sequences which encode secreted or transmembrane PRO	
CC	polypeptides. The PRO polynucleotides and polypeptides have cytosolic	
CC	activity. The polynucleotides and polypeptides can be used for detecting	
CC	the presence of PRO polypeptides in samples, for linking bioactive	
CC	molecules to cells and for modulating biological activities of cells,	
CC	using the polypeptides for specific targeting. The polypeptide targeting	
CC	can be used to kill the target cells, e.g. for the treatment of cancers.	
CC	The polypeptide pairs provide specific targeting of bioactive molecules	
CC	to cells. AAC78600 to AAC78987 represent PCR primers and probes used in	
CC	the isolation of the PRO polynucleotide sequences.	
XX		
SQ	Sequence 67 AA:	
	Query Match 0.7%; Score 8; DB:21; Length 67;	
	Best Local Similarity 100.0%; Pred. No. 24;	
	Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	1133 GLLLALL 1140 	
Db	12 GLTLLALL 19	
RESULT 70		
AAB19581		
ID	AAB19581 standard; Protein: 67 AA.	
XX		
XC	AAB19581;	
XX		
DT	22-JAN-2001 (first entry)	
DE		
XX	Human PRO617 used to treat ocular diseasee.	
XX		
KW	PRO617; human; ocular disease; retinopathy; maculopathy; therapy;	
KW	retinitis pigmentosa; macular degeneration; retinal detachment;	
KW	retinal tear; macular hole; myopia; traumatic choriorretinopathy;	
KW	acute retinal necrosis syndrome; conjunctiva; edema;	
KW	retinal vision occlusion; vascular disease; retinal vasculitis;	
KW	thrombocytopenic purpura; uveitis; retinal occlusion.	
XX		
OS	Homo sapiens.	
XX		
FH	Key Location/Qualifiers	
FT	Peptide 15..27	
FT	/label= Signal_peptide	
FT	Protein 28..67	

FT 1
 FT Modified-site /label= Mature_Pro617
 FT 41.43
 FT /note= "O-phosphorylated"
 XX
 XX MO200053760-A2.
 XX
 XX 14-SEP-2000.
 XX
 XX 10-MAR-2000; 2000WO-US06319.
 XX
 XX 12-MAR-1999; 99US-0123957.
 XX
 XX (GENTH) GENENTECH INC.
 XX
 XX Ferrara N, Goddard A, Gurney AL, Hebert C, Henzel WJ, Kabakoff RC;
 XX Klein RD, Kljavin IJ, Kuo SS, La Fleur M, Wood WI,
 XX MPI: 2000-587437/55.
 XX N-PSDB; AAA88518.
 XX
 XX Novel PRO polypeptides useful for preventing or rescuing retinal cells
 XX from injury caused by ocular diseases such as retinitis pigmentosa,
 XX retinopathy, retinal degenerative diseases, degenerative myopia,
 XX uveitis -
 XX
 XX Claim 2; Fig 10; 140pp; English.
 XX
 XX The present sequence is that of human PRO617, as predicted from a
 XX cDNA clone (see AAA88518) isolated from a foetal kidney tissue cDNA
 XX library using probes and primers (see AAA88535-37) based on a cDNA
 XX clone isolated from an expressed sequence tag database screening
 XX using about 950 secreted protein extracellular domains. PRO617
 XX has a predicted mol.wt. of 6,981 and a pI of about 7.47. A
 XX method for producing PRO polypeptides, including PRO617, using a
 XX host cell transformed with a vector comprising a PRO nucleic acid is
 XX claimed. The invention relates to the use of PRO polypeptides to delay,
 XX prevent or rescue retinal cells such as retinal neurons selected from
 XX photoreceptors, retinal ganglion cells, displaced retinal ganglion
 XX cells, amacrine cells, displaced amacrine cells, horizontal and
 XX bipolar neurons, and supportive cells (including Muller cells and
 XX pigment epithelial cells) from injury and degradation. The retinal
 XX cells are preferably photoreceptors and photoreceptor cell injury or
 XX death is caused by retinal injury, light or environmental trauma or
 XX by an ocular disease selected from retinitis pigmentosa, macular
 XX degeneration, including age-related, retinal detachment, retinal
 XX tears, retinopathy, retinal degenerative diseases, macular holes,
 XX degenerative myopia, acute retinal necrosis syndrome, traumatic
 XX chorioretinopathies or contusion such as Purtscher's retinopathy,
 XX edema, ischemic conditions such as central or branch retinal vision
 XX occlusion, collagen vascular diseases, thrombocytopenic purpura,
 XX uveitis, retinal vasculitis and occlusion associated with Eales
 XX disease and systemic lupus erythematosus (claimed).
 XX
 XX Sequence 67 AA:
 XX
 XX Query Match 0.7%; Score 8; DB 21; Length 67;
 XX Best Local Similarity 100.0%; Pred. No. 24;
 XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1133 GLLALL 1140
 XX |||||
 XX DB 12 GLLALL 19
 XX
 XX RESULT 71
 XX AAU28155
 XX ID AAU28155 standard; Protein; 67 AA.
 XX
 XX AC AAU28155;
 XX
 XX DT 18-DEC-2001 (first entry)
 XX
 XX DE Novel human secretory protein, Seq ID No 324.

XX
 XX Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
 XX ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
 XX transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
 XX amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
 XX ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
 XX gut protection; lung; liver fibrosis; immune deficiency; infection;
 XX severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
 XX multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
 XX fertility; analgesic; pain; antigen.
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO20016689-A2.
 XX
 XX PD 13-SEP-2001.
 XX
 XX PF 05-MAR-2001; 2001WO-US04942.
 XX
 XX PR 07-MAR-2000; 2000US-0519705.
 XX PR 19-MAY-2000; 2000US-0574454.
 XX PR 17-JUN-2000; 2000US-0596193.
 XX PR 14-JUL-2000; 2000US-0616847.
 XX PR 19-SEP-2000; 2000US-0665363.
 XX PR 20-OCT-2000; 2000US-0693267.
 XX
 XX PA (HYSE-) HYSEQ INC.
 XX
 XX PI Tang YT, Liu C, Asundi V, Xu C, Wehman T, Ren F, Ma Y, Zhou P;
 XX PI Zhao Q, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AD, Wang J;
 XX MPI: 2001-589934/66.
 XX N-PSDB; AAS45055.
 XX
 XX PT Novel polypeptides and nucleic acids obtained from cDNA libraries
 XX prepared from various human tissues, for diagnosis and treatment of
 XX cancer, neurological, inflammatory, and autoimmune disorders -
 XX
 XX Example 4; SEQ ID No 324; 107pp; English.
 XX
 XX The invention relates to novel isolated human secreted polypeptides (I)
 XX and polynucleotides (II). (I) and (II) are useful for treating
 XX inflammatory conditions such as arthritis, nephritis, Crohn's disease,
 XX ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
 XX involved in increasing haematopoiesis, stem cell survival, bone growth
 XX and remodeling. (I), (II) and modulators of (II) are useful for
 XX prophylaxis or treatment of one or more cancers. (II) is also useful for
 XX creating transgenic animals useful for studying the in vivo activities of
 XX the polypeptide as well as for studying modulators of the polypeptides.
 XX (I) induces the proliferation of neural cells and regeneration of nerve
 XX and brain tissue and is useful for the treatment of central and
 XX peripheral nervous system diseases and neuropathies, such as Alzheimer's,
 XX Parkinson's disease, Huntington's disease, and amyotrophic lateral
 XX sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
 XX activity, regulation of haematopoiesis and is useful for treating myeloid
 XX or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 XX and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 XX tissue growth, and in tissue repair, healing of burns, incisions,
 XX ulcers, for treating osteoporosis, osteoarthritis, bone degenerative
 XX disorders, or periodontal disease. Furthermore, (I) is also useful for
 XX gut protection or regeneration and treatment of lung or liver fibrosis,
 XX reperfusion injury in various tissues, various immune deficiencies and
 XX disorders including severe combined immunodeficiency (SCID), bacterial or
 XX fungal infections, autoimmune disorders e.g. multiple sclerosis,
 XX rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
 XX reactions and conditions, such as asthma or other respiratory problems.
 XX In addition, (I) affects biorhythms or circadian cycles of rhythms,
 XX fertility, metabolism, catabolism, anabolism, storage or elimination of
 XX dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
 XX analgesic effects or other pain reducing effects, immunoglobulin like
 XX activity and can act as an antigen in a vaccine composition to raise an
 XX immune response. AAU28020-AAU28395 represent novel human secreted protein
 XX amino acid sequences of the invention.

XX Sequence 67 AA; Score 8; DB 22; Length 67;
SQ Query Match Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1133 GLLLLALL 1140
12 GLLLLALL 19
Db

RESULT 72
AAU74938
ID AAU74938 standard; Protein; 67 AA.
XX AAU74938;
XX
XX 23-APR-2002 (first entry)
XX
XX Human clone DNA48309-1280 amino acid sequence of PRO617 protein.
DE
XX Human; clone DNA48309-1280; PRO617; immune response modulator;
XX PRO polypeptide; immune related disorder; systemic lupus erythematosus;
XX rheumatoid arthritis; osteoarthritis; systemic sclerosis; sarcoidosis;
XX autoimmune thrombocytopenia; thyroiditis; diabetes mellitus; anaemia;
XX multiple sclerosis; inflammatory bowel disease; allergic rhinitis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 15..27
FT /label= Signal_peptide
FT 28..67
FT Protein /label= Mature_PRO617_protein
XX
XX WO200192331-A2.
XX
XX 06-DEC-2001.
XX
XX 25-MAY-2001; 2001WO-US17092.
XX
XX 30-MAY-2000; 2000WO-US14941.
XX 02-JUN-2000; 2000WO-US15264.
XX 01-DEC-2000; 2000WO-US32678.
XX
XX (GENTH) GENENTECH INC.
XX
XX Eaton DL, Fong S, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL;
XX Watanabe CK, Wood WI, Zhang Z;
XX
XX MPI: 2002-075461/10.
XX N-PSDB; ABK14010.
XX
XX Isolated PRO221, PRO617, PRO1030, PRO4302 polypeptides, useful for
XX creating immune disorders such as thyroiditis, diabetes mellitus,
XX allergic disease, asthma, allergic rhinitis, atopic dermatitis -
XX
XX Claim 10; Fig 4; 104dp; English.
XX
XX The present invention relates to a new polypeptide having at least 80%
XX amino acid sequence identity to fully defined amino acid sequence of
XX PRO221 (AAU74937), PRO617 (AAU74938), PRO1030 (AAU74939) or PRO4302
XX (AAU74940) as given in specification. The PRO polypeptides of the
XX invention are useful for treating an immune related disorder such as
XX systemic lupus erythematosus, rheumatoid arthritis, osteoarthritis,
XX systemic sclerosis, sarcoidosis, autoimmune thrombocytopenia,
XX thyroiditis, diabetes mellitus, multiple sclerosis, inflammatory bowel
XX disease, allergic rhinitis and anaemia. The methods of the invention
XX are useful for identifying an antisense nucleic acid and for enhancing
XX the infiltration of inflammatory cells into the tissue of a mammal. The
XX anti-PRO antibodies are useful for determining the presence of PRO
XX polypeptides and for diagnosing an immune-related disease in a mammal.

CC The molecules of the invention are useful for identifying a compound
CC that inhibits the activity of PRO polypeptides or expression of a gene
CC encoding the PRO polypeptide. PRO polypeptides are also useful for
CC increasing or decreasing the infiltration of inflammatory cells from the
CC vasculature into a tissue of a mammal, increasing or decreasing activity
CC or proliferation of T-lymphocytes. The present amino acid sequence
CC represents the PRO617 polypeptide of the invention. The PRO617 sequence
CC was isolated from the clone DNA48309-1280.
XX
XX SQ Sequence 67 AA; Score 8; DB 23; Length 67;
Query Match Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1133 GLLLLALL 1140
12 GLLLLALL 19
Db

RESULT 73
ABB17942
ID ABB17942 standard; Protein; 69 AA.
XX ABB17942;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polypeptide SEQ ID NO 6599.
DE
XX Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
XX antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
XX antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 HFLPLVLF 15
 |||||
 Db 23 HFLPLVLF 30

RESULT 74
 AAB47587
 ID AAB47587 standard; peptide; 76 AA.
 XX
 AC AAB47587;
 XX
 DT 13-DEC-2001 (first entry)
 XX
 DE Mouse heat shock antigen peptide.
 XX
 KW Mouse; heat shock antigen; HSA; human; rat; signal transducer; CD24;
 KW fusion protein; inhibition; autoreactive T cell; atc;
 KW autoimmune disease; multiple sclerosis; rheumatoid arthritis;
 KW systemic lupus erythematosus; psoriasis; diabetes; allergy;
 KW transplant rejection; transgenic mouse.
 XX
 OS Mus musculus.
 XX

Key Location/Qualifiers
 FH Peptide 1..26
 FT /label= Signal peptide
 FT Modified-site 27
 FT /label= Glycosylation site
 FT Protein 27..53
 FT /label= Mature signal transducer CD24
 FT Modified-site 30
 FT /label= Glycosylation site
 FT Modified-site 39
 FT /label= Glycosylation site
 FT Modified-site 48
 FT /label= Glycosylation site
 FT Binding-site 53
 FT /label= Lipid binding site
 FT /note= "Potential GPI anchor"
 FT 54..76
 FT /label= Pro-peptide
 FT /note= "removed in mature form"

MO200172325-A1.
 XX
 PD 04-OCT-2001.
 XX
 PF 29-MAR-2001; 2001MO-US40390.
 XX
 PR 29-MAR-2000; 2000US-192814P.
 XX
 PA (OHIS) UNIV OHIO STATE RES FOUND.
 XX
 PI Liu Y, Zheng P, Bai X;
 XX
 DR WPI; 2001-611581/70.
 XX
 PT Inhibiting tissue destruction by autoreactive T cells, useful for
 PT creating autoimmune diseases, by administering a heat-shock
 PT antigen/CD24 polypeptide or its antibody -
 XX
 PS Disclosure; Fig 7; 34pp; English.
 XX
 CC The sequences given in AAB47587-89 represent mouse heat shock antigen
 CC (HSA) and its homologs in human and rat, signal transducer CD24
 CC protein. These proteins may be used, in a composition or in a fusion
 CC protein, in the method of the invention for inhibiting destruction of
 CC tissue initiated by autoreactive T cells (atc). The method is
 CC especially used to treat subjects suspected of having autoimmune
 CC diseases, particularly multiple sclerosis, rheumatoid arthritis,
 CC systemic lupus erythematosus, psoriasis, diabetes and allergy, also

CC transplant rejection. Transgenic mice that express human CD24 on their
 CC T cells are useful as models for testing drugs for use against
 CC autoimmune diseases.
 XX

QY 1133 GLLLLALL 1140
 |||||
 Db 12 GLLLLALL 19

RESULT 75
 AAB47589
 ID AAB47589 standard; peptide; 76 AA.
 XX
 AC AAB47589;
 XX
 DT 13-DEC-2001 (first entry)
 XX
 DE Rat CD24 peptide.
 XX
 KW Mouse; heat shock antigen; HSA; human; rat; signal transducer; CD24;
 KW fusion protein; inhibition; autoreactive T cell; atc;
 KW autoimmune disease; multiple sclerosis; rheumatoid arthritis;
 KW systemic lupus erythematosus; psoriasis; diabetes; allergy;
 KW transplant rejection; transgenic mouse.
 XX
 OS Rattus rattus.
 XX

Key Location/Qualifiers
 FH Peptide 1..26
 FT /label= Signal peptide
 FT Domain 57..76
 FT /label= GPI anchor region

MO200172325-A1.
 XX
 PD 04-OCT-2001.
 XX
 PF 29-MAR-2001; 2001MO-US40390.
 XX
 PR 29-MAR-2000; 2000US-192814P.
 XX
 PA (OHIS) UNIV OHIO STATE RES FOUND.
 XX
 PI Liu Y, Zheng P, Bai X;
 XX
 DR WPI; 2001-611581/70.
 XX
 PT Inhibiting tissue destruction by autoreactive T cells, useful for
 PT creating autoimmune diseases, by administering a heat-shock
 PT antigen/CD24 polypeptide or its antibody -
 XX
 PS Disclosure; Fig 9; 34pp; English.
 XX
 CC The sequences given in AAB47587-89 represent mouse heat shock antigen
 CC (HSA) and its homologs in human and rat, signal transducer CD24
 CC protein. These proteins may be used, in a composition or in a fusion
 CC protein, in the method of the invention for inhibiting destruction of
 CC tissue initiated by autoreactive T cells (atc). The method is
 CC especially used to treat subjects suspected of having autoimmune
 CC diseases, particularly multiple sclerosis, rheumatoid arthritis,
 CC systemic lupus erythematosus, psoriasis, diabetes and allergy, also
 CC transplant rejection. Transgenic mice that express human CD24 on their
 CC T cells are useful as models for testing drugs for use against
 CC autoimmune diseases.
 XX

SQ Sequence 76 AA;

Query Match 0.7%; Score 8; DB 22; length 76;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1133 GLLTLALL 1140
Db 12 GLLTLALL 19

Search completed: July 16, 2003, 08:32:18
Job time : 52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 08:13:03 ; Search time 14 Seconds
(without alignments)
46.236 Million cell updates/sec

Title: US-09-647-544-7
Sequence: 1 KLGFPAHKKIPPEEKREKLEQ 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 140335

Minimum DB seq length: 0
Maximum DB seq length: 22

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents, AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill.esl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	37.7	20	1 US-08-214-770-1	Sequence 1, Appli
2	43	37.7	20	5 PCT-US95-02885-1	Sequence 1, Appli
3	43	37.7	20	5 PCT-US95-07542-15	Sequence 15, Appli
4	39	34.2	20	6 5225193-7	Patent No. 5225193
5	36	31.6	20	6 5225193-2	Patent No. 5225193
6	36	31.6	21	1 US-08-073-028-25	Sequence 25, Appli
7	36	31.6	21	1 US-08-127-499A-15	Sequence 15, Appli
8	36	31.6	21	1 US-08-482-847-15	Sequence 15, Appli
9	36	31.6	21	4 US-08-554-616-25	Sequence 25, Appli
10	36	31.6	22	1 US-08-073-028-24	Sequence 24, Appli
11	36	31.6	22	4 US-08-554-616-24	Sequence 24, Appli
12	35	30.7	21	5 PCT-US95-07542-11	Sequence 11, Appli
13	34	29.8	13	1 US-08-056-200-11	Sequence 11, Appli
14	34	29.8	13	1 US-07-987-272A-6	Sequence 6, Appli
15	34	29.8	13	2 US-08-800-644-11	Sequence 11, Appli
16	34	29.8	18	2 US-08-649-991-49	Sequence 49, Appli
17	34	29.8	19	2 US-08-690-011A-4	Sequence 4, Appli
18	34	29.8	19	4 US-09-299-495F-4	Sequence 4, Appli
19	34	29.8	20	1 US-08-199-776-21	Sequence 21, Appli
20	34	29.8	20	3 US-08-663-731-21	Sequence 21, Appli
21	34	29.8	20	3 US-08-879-338-21	Sequence 21, Appli
22	34	29.8	20	4 US-09-406-781-1	Sequence 1, Appli
23	34	29.8	20	4 US-09-181-941-15	Sequence 15, Appli
24	34	29.8	20	4 US-09-293-238B-21	Sequence 21, Appli
25	34	29.8	20	5 PCT-US95-02044-21	Sequence 21, Appli
26	34	29.8	21	1 US-08-199-776-5	Sequence 5, Appli
27	34	29.8	21	1 US-08-555-579-1	Sequence 1, Appli

28	34	29.8	21	3 US-08-663-731-5	Sequence 5, Appli
29	34	29.8	21	3 US-08-879-338-5	Sequence 5, Appli
30	34	29.8	21	3 US-08-879-338-26	Sequence 26, Appli
31	34	29.8	21	4 US-09-293-238B-5	Sequence 5, Appli
32	34	29.8	21	4 US-09-293-238B-26	Sequence 26, Appli
33	34	29.8	21	5 PCT-US95-02044-5	Sequence 5, Appli
34	33.5	29.4	22	6 5104631-11	Patent No. 5104631
35	33	28.9	14	1 US-08-477-108A-10	Sequence 10, Appli
36	33	28.9	15	1 US-08-199-776-20	Sequence 20, Appli
37	33	28.9	15	3 US-08-663-731-20	Sequence 20, Appli
38	33	28.9	15	3 US-08-879-338-20	Sequence 20, Appli
39	33	28.9	15	4 US-09-293-238B-20	Sequence 20, Appli
40	33	28.9	15	5 PCT-US95-02044-20	Sequence 20, Appli
41	33	28.9	21	4 US-08-584-043A-81	Sequence 81, Appli
42	32	28.1	16	4 US-09-120-653D-9	Sequence 9, Appli
43	32	28.1	18	1 US-08-351-423-3	Sequence 3, Appli
44	32	28.1	18	3 US-08-940-095-249	Sequence 249, App
45	32	28.1	18	3 US-08-940-093-249	Sequence 249, App

ALIGNMENTS

RESULT 1
US-08-214-770-1
; Sequence 1, Application US/08214770
; Patent No. 5523209
; GENERAL INFORMATION:
; APPLICANT: Ginsberg, Mark H.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: INHIBITORS OF INTEGRIN
; TITLE OF INVENTION: ACTIVATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,770
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06410/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-214-770-1

Query Match 37.7%; Score 43; DB 1; Length 20;
Best Local Similarity 50.0%; Pred. NO. 6.8;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Mathews, Thomas J.
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds,
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,028
FILING DATE: 07-JUN-1993
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-004-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-073-028-25

Query Match 31.6%; Score 36; DB 1; Length 21;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 7 HKKIPKREKREKLEQ 22
Db 6 HSLIESQNGQCKNEQ 21

RESULT 7
US-08-127-499A-15
Sequence 15, Application US/08127499A
Patent No. 5510264
GENERAL INFORMATION:
APPLICANT: VAN ALSTYNE, Diane
APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
TITLE OF INVENTION: HOMOLOGOUS ANTIGENIC SEQUENCES
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,499A
FILING DATE: 28-SEP-1993

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/102/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-127-499A-15

Query Match 31.6%; Score 36; DB 1; Length 21;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 7 HKKIPKREKREKLEQ 22
Db 1 HSLIESQNGQCKNEQ 16

RESULT 8
US-08-482-847-15
Sequence 15, Application US/08482847
Patent No. 5556757
GENERAL INFORMATION:
APPLICANT: VAN ALSTYNE, Diane
APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THERETO, AND USES THEREOF
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,847
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/127,499
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/104/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-482-847-15

Query Match 31.6%; Score 36; DB 1; Length 21;
Best Local Similarity 43.8%; Pred. No. 65;

	Matches	7; Conservative	3; Mismatches	6; Indels	0; Gaps	0
QY	7	HKKIPBEKKREKLEQ	22			
Db	1	HSLEESQNGQEKNEQ	16			

```

RESULT 9
US-08-554-616-25
Sequence 25, Application US/08554616
Patent No. 6133418
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/554,616
FILING DATE: 06-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,028
FILING DATE: 07-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30, 742
REFERENCE/DOCKET NUMBER: 7872-004-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-554-616-25

Query Match 31.6%; Score 36; DB 4; Length 21;
Best Local Similarity 43.8%; Pred. No. 65;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0.

OY 7 HKKIPKREKLEQ 22
DB 6 HSLIESQNOQEKNEQ 21

RESULT 10
US-08-073-028-24
Sequence 24, Application US/08073028
Patent No. 5464933
GENERAL INFORMATION:
APPLICANT: Bolognesi, Dani P.
APPLICANT: Matthews, Thomas J.
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
TITLE OF INVENTION: TRANSMISSION
NUMBER OF SEQUENCES: 74

```

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7872-004-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-554-616-24

Query Match 31.6%; Score 36; DB 4; Length 22;
Best Local Similarity 43.8%; Pred. No. 68;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 7 HKIPEEKREKLEQ 22
Db 6 HSLIESONQOEKNEQ 21

RESULT 12
PCT-US95-07542-11
Sequence 11, Application PC/TUS9507542
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: STRUCTURAL MODELS FOR CYTOPLASMIC
TITLE OF INVENTION: DOMAINS OF TRANSMEMBRANE RECEPTORS
NUMBER OF SEQUENCES: 20
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07542
FILING DATE: 13-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/260,514
FILING DATE: 15-JUN-1994
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: C-terminal
ORIGINAL SOURCE:
ORGANISM: Synthetic peptide modeling integrin region
PCT-US95-07542-11

Query Match 30.7%; Score 35; DB 5; Length 21;
Best Local Similarity 44.4%; Pred. No. 88;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 KLGFPAAKKIPEEKREE 18
Db 2 KVGFPKRRHATLEEDDEE 19

RESULT 13
US-08-056-200-11
Sequence 11, Application US/08056200
GENERAL INFORMATION:
APPLICANT: Steiner, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESS: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-056-200-11

Query Match 29.8%; Score 34; DB 1; Length 13;
Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 12 EEEKREKXL 20
Db 5 EEEKREKQL 13

RESULT 14
US-07-987-272A-6
Sequence 6, Application US/07987272A
Patent No. 5731166
GENERAL INFORMATION:
APPLICANT: Geary, C., Simpson, R. J. and Lackmann, M
TITLE OF INVENTION: No. 5731166el Chemotactic Factor
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Cushman Darby & Cushman
STREET: 1100 New York Avenue, N. W., Ninth Floor, East Tower
CITY: Washington
STATE: D. C.
COUNTRY: USA
ZIP: 20005-3918

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/987,272A
FILING DATE: 05-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 2127
FILING DATE: 05-FEB-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK 4463
FILING DATE: 05-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Brinkman, David W
REGISTRATION NUMBER: 20,817
REFERENCE/DOCKET NUMBER: DWB/1925/200259
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861 3000
TELEFAX: 202-822 0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-987-272A-6

Query Match 29.8%; Score 34; DB 1; Length 13;
Best Local Similarity 58.3%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 AHKKIPEEEKRE 17
DB 2 AHKKSHEESHKE 13

RESULT 15
US-08-800-644-11
Sequence 11, Application US/08800644
Patent No. 5958752

GENERAL INFORMATION:

APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Seo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,644
FILING DATE: 14-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/056,200
FILING DATE: 30-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE:
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-800-644-11

Query Match 29.8%; Score 34; DB 2; Length 13;
Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 12 EEEKREKRL 20
DB 5 EEEKREKRL 13

Search completed: July 16, 2003, 08:15:02
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 16, 2003, 08:31:25 ; Search time 18 Seconds
(without alignments)
1907.586 Million cell updates/sec

Title: US-09-647-544-2

Perfect score: 1167
Sequence: 1 MELPFVTHLPLVFLTGLC.....GFFAHKKIPEEKREKLEQ 1167

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	0.9	1178	1	US-08-199-776-2
2	10	0.9	1178	3	US-08-663-731-2
3	10	0.9	1178	3	US-08-879-338-2
4	10	0.9	1178	5	PCT-US95-02044-2
5	10	0.9	1178	4	US-09-293-238B-2
6	9	0.8	1065	3	US-08-630-172-9
7	9	0.8	1065	4	US-09-375-419-9
8	9	0.8	1170	2	US-08-789-078-2
9	9	0.8	1170	2	US-08-476-633-2
10	9	0.8	1170	2	US-08-476-062A-42
11	9	0.8	1170	5	PCT-US95-04886-2
12	9	0.8	1170	5	PCT-US96-01314-42
13	8	0.7	17	3	US-08-893-526A-29
14	8	0.7	76	1	US-08-848-525-4
15	8	0.7	76	3	US-09-083-521-5
16	8	0.7	80	1	US-08-848-525-2
17	8	0.7	80	1	US-08-848-525-2
18	8	0.7	104	4	US-08-973-544-4
19	8	0.7	104	4	US-08-973-544-3
20	8	0.7	179	4	US-09-419-568F-28
21	8	0.7	234	1	US-09-354-243B-28
22	8	0.7	494	1	US-08-485-618-103
23	8	0.7	494	2	US-08-605-672-37
24	8	0.7	494	2	US-08-482-293A-103
25	8	0.7	494	2	US-08-943-363-103
26	8	0.7	494	2	US-09-193-043-103
27	8	0.7	494	4	US-09-688-307A-103

28	8	0.7	513	4	US-09-291-922-20	Sequence 20, Appli
29	8	0.7	1039	4	US-09-409-648-7	Sequence 7, Appli
30	8	0.7	1039	4	US-09-409-648-8	Sequence 8, Appli
31	8	0.7	1039	6	5196511-2	Patent No. 5196511
32	8	0.7	1151	1	US-08-286-889-37	Sequence 37, Appli
33	8	0.7	1151	1	US-08-485-618-37	Sequence 37, Appli
34	8	0.7	1151	1	US-08-362-652-37	Sequence 37, Appli
35	8	0.7	1151	2	US-08-605-672-37	Sequence 37, Appli
36	8	0.7	1151	2	US-08-482-293A-37	Sequence 37, Appli
37	8	0.7	1151	2	US-08-943-363-37	Sequence 37, Appli
38	8	0.7	1151	4	US-09-193-043-37	Sequence 37, Appli
39	8	0.7	1151	4	US-09-688-307A-37	Sequence 37, Appli
40	8	0.7	1152	2	US-08-476-062A-43	Sequence 43, Appli
41	8	0.7	1152	5	PCT-US96-01314-43	Sequence 43, Appli
42	8	0.7	1152	6	5424399-2	Patent No. 5424399
43	8	0.7	1153	1	US-08-173-497-3	Sequence 3, Appli
44	8	0.7	1153	1	US-08-286-889-3	Sequence 3, Appli
45	8	0.7	1153	1	US-08-485-618-3	Sequence 3, Appli
46	8	0.7	1153	1	US-08-362-652-3	Sequence 3, Appli
47	8	0.7	1153	2	US-08-605-672-3	Sequence 3, Appli
48	8	0.7	1153	2	US-08-482-293A-3	Sequence 3, Appli
49	8	0.7	1153	2	US-08-943-363-3	Sequence 3, Appli
50	8	0.7	1153	4	US-09-193-043-3	Sequence 3, Appli
51	8	0.7	1153	4	US-09-688-307A-3	Sequence 3, Appli
52	8	0.7	1155	1	US-08-286-889-46	Sequence 46, Appli
53	8	0.7	1155	1	US-08-485-618-46	Sequence 46, Appli
54	8	0.7	1155	1	US-08-362-652-46	Sequence 46, Appli
55	8	0.7	1155	2	US-08-605-672-46	Sequence 46, Appli
56	8	0.7	1155	2	US-08-482-293A-46	Sequence 46, Appli
57	8	0.7	1155	2	US-08-943-363-46	Sequence 46, Appli
58	8	0.7	1155	4	US-09-193-043-46	Sequence 46, Appli
59	8	0.7	1155	4	US-09-688-307A-46	Sequence 46, Appli
60	8	0.7	1161	1	US-08-485-618-53	Sequence 53, Appli
61	8	0.7	1161	1	US-08-485-618-55	Sequence 53, Appli
62	8	0.7	1161	1	US-08-362-652-53	Sequence 53, Appli
63	8	0.7	1161	1	US-08-362-652-55	Sequence 53, Appli
64	8	0.7	1161	2	US-08-605-672-53	Sequence 53, Appli
65	8	0.7	1161	2	US-08-605-672-55	Sequence 53, Appli
66	8	0.7	1161	2	US-08-482-293A-53	Sequence 53, Appli
67	8	0.7	1161	2	US-08-943-363-53	Sequence 53, Appli
68	8	0.7	1161	2	US-08-943-363-55	Sequence 53, Appli
69	8	0.7	1161	2	US-08-943-363-55	Sequence 53, Appli
70	8	0.7	1161	4	US-09-193-043-53	Sequence 53, Appli
71	8	0.7	1161	4	US-09-193-043-55	Sequence 53, Appli
72	8	0.7	1161	4	US-09-688-307A-53	Sequence 53, Appli
73	8	0.7	1161	4	US-09-688-307A-55	Sequence 53, Appli
74	8	0.7	1163	1	US-08-173-497-4	Sequence 4, Appli
75	8	0.7	1163	1	US-08-286-889-4	Sequence 4, Appli
76	8	0.7	1163	1	US-08-485-618-4	Sequence 4, Appli
77	8	0.7	1163	1	US-08-362-652-4	Sequence 4, Appli
78	8	0.7	1163	2	US-08-605-672-4	Sequence 4, Appli
79	8	0.7	1163	2	US-08-482-293A-4	Sequence 4, Appli
80	8	0.7	1163	2	US-08-943-363-4	Sequence 4, Appli
81	8	0.7	1163	2	US-08-943-363-4	Sequence 4, Appli
82	8	0.7	1163	4	US-08-476-062A-44	Sequence 44, Appli
83	8	0.7	1163	4	US-09-193-043-4	Sequence 4, Appli
84	8	0.7	1163	5	PCT-US96-01314-44	Sequence 44, Appli
85	8	0.7	1163	5	PCT-US96-01314-44	Sequence 44, Appli
86	8	0.7	1580	2	US-08-804-227C-11	Sequence 11, Appli
87	8	0.7	1580	2	US-08-804-198-5	Sequence 11, Appli
88	8	0.7	1891	2	US-08-804-227C-12	Sequence 12, Appli
89	8	0.7	1891	2	US-08-804-198-6	Sequence 12, Appli
90	7	0.6	9	1	US-08-463-128-4	Sequence 34, Appli
91	7	0.6	16	1	US-08-463-180-34	Sequence 34, Appli
92	7	0.6	16	1	US-08-488-212A-33	Sequence 33, Appli
93	7	0.6	16	2	US-08-320-306-33	Sequence 33, Appli
94	7	0.6	16	2	US-08-488-209B-33	Sequence 33, Appli
95	7	0.6	17	3	US-08-408-011-33	Sequence 33, Appli
96	7	0.6	68	1	US-08-893-526A-21	Sequence 21, Appli
97	7	0.6	95	1	US-07-756-250-10	Sequence 10, Appli
98	7	0.6	95	1	US-07-626-618A-2	Sequence 2, Appli
99	7	0.6	95	1	US-07-928-611-2	Sequence 2, Appli
100	7	0.6	95	1	US-08-333-977-2	Sequence 2, Appli
					US-08-487-811A-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-199-776-2
; Sequence 2, Application US/08199776
; Patent No. 5594120
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: No. 5594120el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,776
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-199-776-2

Query Match          0.9%; Score 10; DB 1; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      483 GSYFGSELCP 492
      |||||
      510 GSYFGSELCP 519

RESULT 2
US-08-663-731-2
; Sequence 2, Application US/08663731
; Patent No. 6057423
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: No. 6057423el integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,731
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/199,776
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1178 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-663-731-2

Query Match          0.9%; Score 10; DB 3; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      483 GSYFGSELCP 492
      |||||
      510 GSYFGSELCP 519

RESULT 3
US-08-879-338-2
; Sequence 2, Application US/08879338A
; Patent No. 6063906
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: Antibodies to No. 6063906el Integrin Alpha
; FILE REFERENCE: B0801/7080/ERP
; CURRENT APPLICATION NUMBER: US/08/879,338A
; EARLIER FILING DATE: 1997-06-20
; EARLIER APPLICATION NUMBER: US 08/663,731
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: US 08/199,776
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1178
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (-18)...(-1)
US-08-879-338-2

Query Match          0.9%; Score 10; DB 3; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      483 GSYFGSELCP 492
      |||||
      510 GSYFGSELCP 519

RESULT 4
PCR-US95-02044-2
; Sequence 2, Application PC/TUS9502044
; GENERAL INFORMATION:
```


APPLICANT:
TITLE OF INVENTION: Novel integrin alpha subunit
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02044
FILING DATE: herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE: 18 February 1994
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-2441
TELEFAX: 617-720-3500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1178 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02044-2

Query Match 0.9%; Score 10; DB 5; Length 1178;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSELCP 492
|||||
Db 510 GSYFSELCP 519

RESULT 5
US-09-293-238B-2
Sequence 2, Application US/09293238B
Patent No. 6455042
GENERAL INFORMATION:
APPLICANT: Brenner, Michael B.
TITLE OF INVENTION: A Method of Treating Ulcerative Colitis
TITLE OF INVENTION: or Crohn's Disease by Administering an Antibody to Alpha E
FILE REFERENCE: L0560/7005/ERP
CURRENT APPLICATION NUMBER: US/09/293,238B
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: US 08/879,338
PRIOR FILING DATE: 1997-06-20
PRIOR APPLICATION NUMBER: US 08/663,731
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: US 08/199,776
PRIOR FILING DATE: 1994-02-18
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1179
TYPE: PRT
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: SIGNAL

LOCATION: (1)...(18)
US-09-293-238B-2

Query Match 0.9%; Score 10; DB 4; Length 1179;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 483 GSYFSELCP 492
|||||
Db 511 GSYFSELCP 520

RESULT 6
US-08-630-172-9
Sequence 9, Application US/08630172
Patent No. 6060054
GENERAL INFORMATION:
APPLICANT: Staetz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,172
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-630-172-9

Query Match 0.8%; Score 9; DB 3; Length 1065;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
|||||
Db 514 DVAVGAPLE 522

RESULT 7
US-09-375-419-9
Sequence 9, Application US/09375419
Patent No. 6264950
GENERAL INFORMATION:
APPLICANT: Staetz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor

CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1065 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-9

Query Match 0.8%; Score 9; DB 4; Length 1065;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
|||||||
DB 514 DVAVGAPLE 522

RESULT 8
US-08-789-078-2
Sequence 2, Application US/08789078
Patent No. 5843885
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbels, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,078
FILING DATE: 03-FEB-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,513
FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.

REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: T-cell
CELL LINE: HL-60
FEATURE:
NAME/KEY: Region
LOCATION: 1..25
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "Signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 32..79
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat I"
FEATURE:
NAME/KEY: Region
LOCATION: 82..132
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat II"
FEATURE:
NAME/KEY: Region
LOCATION: 339..391
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat III"
FEATURE:
NAME/KEY: Region
LOCATION: 392..446
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat IV"
FEATURE:
NAME/KEY: Region
LOCATION: 447..508
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat V"
FEATURE:
NAME/KEY: Region
LOCATION: 509..567
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat VI"
FEATURE:
NAME/KEY: Region
LOCATION: 568..629
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat VII"
FEATURE:
NAME/KEY: Domain
LOCATION: 170..349
OTHER INFORMATION: /label= IDomain
OTHER INFORMATION: /note= "I-Domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 1089..1112
OTHER INFORMATION: /label= Trans
OTHER INFORMATION: /note= "Transmembrane Domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 1113..1170
OTHER INFORMATION: /label= Cyto
OTHER INFORMATION: /note= "Cytoplasmic domain"
PUBLICATION INFORMATION:

AUTHORS: Pigott,
ANTHORS: Power,
TITLE: LFA-1 Amino acid sequence (alpha) (from
TITLE: PMA-stimulated HL-60 cells)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 94-95
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170
US-08-789-078-2

Query Match 0.8%; Score 9; DB 2; Length 1170;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
|||
Db 538 DVAVGAPLE 546

RESULT 9
US-08-752-633-2
Sequence 2, Application US/08752633
Patent No. 5863889
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Sahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbets, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: T-cell
CELL LINE: HL-60
FEATURE:
NAME/KEY: Region
LOCATION: 1..25
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region

LOCATION: 32..79 /label= Repeat
OTHER INFORMATION: /note= "Repeat I"
FEATURE:
NAME/KEY: Region
LOCATION: 82..132 /label= Repeat
OTHER INFORMATION: /note= "Repeat II"
FEATURE:
NAME/KEY: Region
LOCATION: 339..391 /label= Repeat
OTHER INFORMATION: /note= "Repeat III"
FEATURE:
NAME/KEY: Region
LOCATION: 392..446 /label= Repeat
OTHER INFORMATION: /note= "Repeat IV"
FEATURE:
NAME/KEY: Region
LOCATION: 447..508 /label= Repeat
OTHER INFORMATION: /note= "Repeat V"
FEATURE:
NAME/KEY: Region
LOCATION: 509..567 /label= Repeat
OTHER INFORMATION: /note= "Repeat VI"
FEATURE:
NAME/KEY: Region
LOCATION: 568..629 /label= Repeat
OTHER INFORMATION: /note= "Repeat VII"
FEATURE:
NAME/KEY: Domain
LOCATION: 170..349 /label= IDomain
OTHER INFORMATION: /note= "I-Domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 1089..1112 /label= Trans
OTHER INFORMATION: /note= "Transmembrane Domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 1113..1170 /label= Cyto
OTHER INFORMATION: /note= "Cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (alpha) (from
TITLE: PMA-stimulated HL-60 cells)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 94-95
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170
US-08-752-633-2

Query Match 0.8%; Score 9; DB 2; Length 1170;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
|||
Db 538 DVAVGAPLE 546

RESULT 10
US-08-476-062A-42
Sequence 42, Application US/08476062A
Patent No. 5877275
GENERAL INFORMATION:

APPLICANT: Arnaout, M. Amin
TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,062A
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/216,081
FILING DATE: 21-MAR-1994
APPLICATION NUMBER: 07/637,830
FILING DATE: 04-JAN-1991
APPLICATION NUMBER: 07/539,842
FILING DATE: 18-JUN-1990
APPLICATION NUMBER: 07/212,573
FILING DATE: 28-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John M.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/068003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-476-062A-42

Query Match 0.8%; Score 9; DB 2; Length 1170;
Best local similarity 100.0%; Pred. No. 7.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 566 DVAVGAPLE 574
|||
Db 538 DVAVGAPLE 546

RESULT 11
PCT-US95-04886-2
Sequence 2, Application PC/US9504886
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Stahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbels, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: 816/474-9057
TELEX: 434-363
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: T-cell
CELL LINE: HL-60
FEATURE:
NAME/KEY: Region
LOCATION: 1..25
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "Signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 32..79
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat I"
FEATURE:
NAME/KEY: Region
LOCATION: 82..132
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat II"
FEATURE:
NAME/KEY: Region
LOCATION: 339..391
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat III"
FEATURE:
NAME/KEY: Region
LOCATION: 392..446
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat IV"
FEATURE:
NAME/KEY: Region
LOCATION: 447..508
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat V"
FEATURE:
NAME/KEY: Region
LOCATION: 509..567
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat VI"
FEATURE:
NAME/KEY: Region
LOCATION: 568..629
OTHER INFORMATION: /label= Repeat
OTHER INFORMATION: /note= "Repeat VII"
FEATURE:
NAME/KEY: Domain
LOCATION: 170..349
OTHER INFORMATION: /label= IDomain
OTHER INFORMATION: /note= "I-Domain"
FEATURE:

NAME/KEY: Domain
LOCATION: 1089..1112
OTHER INFORMATION: /label= Trans
OTHER INFORMATION: /note= "Transmembrane Domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 1113..1170
OTHER INFORMATION: /label= Cyto
OTHER INFORMATION: /note= "Cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (alphaL) (from
TITLE: PMA-stimulated HL-60 cells)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 94-95
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 1170
PCT-US95-04886-2

Query Match 0.8%; Score 9; DB 5; Length 1170;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 566 DVAVGAPLE 574
|||||||
DB 538 DVAVGAPLE 546

RESULT 12
PCT-US96-01314-42
Sequence 42, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaout
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1170
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-01314-42

Query Match 0.8%; Score 9; DB 5; Length 1170;
Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 566 DVAVGAPLE 574
|||||||
DB 538 DVAVGAPLE 546

RESULT 13
US-08-893-526A-29
Sequence 29, Application US/08893526A
Patent No. 6096707
GENERAL INFORMATION:
APPLICANT: Heino, Jyrki
APPLICANT: Iwasaka, Joanna
APPLICANT: K pyl, Jarmo
TITLE OF INVENTION: Integrin Binding Peptide and Use Thereof
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave., N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/893,526A
FILING DATE: 11-JUL-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KIM, JUDITH U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 1708, 0240000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/371-2600
TELEFAX: 202/371-2540
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-893-526A-29

Query Match 0.7%; Score 8; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 275 TDGESHG 282
|||||||
DB 1 TDGESHG 8

RESULT 14
US-08-848-252-4
Sequence 4, Application US/08848252
Patent No. 5804177
GENERAL INFORMATION:
APPLICANT: Humphries, Keith R.
TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/848,252
FILING DATE: 29-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/538,052
FILING DATE:
APPLICATION NUMBER: US 08/151,672
FILING DATE: 15-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: McDiarmid, Shona S.
REGISTRATION NUMBER: P-38,798
REFERENCE/DOCKET NUMBER: 3158-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
TELEX: 06-23115
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-848-252-4

Query Match
Best Local Similarity 0.7%; Score 8; DB 1; Length 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
DB 12 GLLALL 19

RESULT 15
US-09-083-521-5
Sequence 5, Application US/09083521
Patent No. 6048970
GENERAL INFORMATION:
APPLICANT: Lai, Preeti
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/083,521
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0527 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1216498
US-09-083-521-5

Query Match
Best Local Similarity 0.7%; Score 8; DB 3; Length 76;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
DB 12 GLLALL 19

RESULT 16
US-08-848-252-2
Sequence 2, Application US/08848252
Patent No. 5804177
GENERAL INFORMATION:
APPLICANT: Humphries, Keith R.
TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/848,252
FILING DATE: 29-APR-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/538,052
FILING DATE:
APPLICATION NUMBER: US 08/151,672
FILING DATE: 15-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: McDiarmid, Shona S.
REGISTRATION NUMBER: P-38,798
REFERENCE/DOCKET NUMBER: 3158-028
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
TELEX: 06-23115
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-848-252-2

Query Match
Best Local Similarity 0.7%; Score 8; DB 1; Length 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1133 GLLALL 1140
DB 12 GLLALL 19

RESULT 17
US-08-973-544-4
Sequence 4, Application US/08973544
Patent No. 6338950
GENERAL INFORMATION:
APPLICANT: WEISS, Elisabeth
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,544
FILING DATE: 18-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT EP 96/02663
FILING DATE: 20-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95109511.6
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95112201.9
FILING DATE: 03-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kites, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P8341-7073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-973-544-4

Query Match 0.7%; Score 8; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138
|||||
DB 16 LGGLLLLA 23

RESULT 18
US-08-973-544-3
Sequence 3, Application US/08973544
Patent No. 6338950
GENERAL INFORMATION:
APPLICANT: WEISS, Elisabeth
TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIKAIIDO, MARCELSTEIN, MURRAY & ORAM LLP
STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-5701

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,544
FILING DATE: 18-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT EP 96/02663
FILING DATE: 20-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95109511.6
FILING DATE: 20-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95112201.9
FILING DATE: 03-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Kites, Monica Chin
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P8341-7073
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 104 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-973-544-3

Query Match 0.7%; Score 8; DB 4; Length 104;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1131 LGGLLLLA 1138
|||||
DB 16 LGGLLLLA 23

RESULT 19
US-09-419-568F-28
Sequence 28, Application US/09419568F
Patent No. 6331613
GENERAL INFORMATION:
APPLICANT: Dumoutier, Laure
TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fac
TITLE OF INVENTION: (Tifs) The Proteins Encoded, and Uses Thereof
FILE REFERENCE: LUD 5543.2
CURRENT APPLICATION NUMBER: US/09/419,568F
CURRENT FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: US09/354,243
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: US09/178,973
PRIOR FILING DATE: 1998-10-26
NUMBER OF SEQ ID NOS: 29
SEQ ID NO 28
LENGTH: 179
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
US-09-419-568F-28

Query Match 0.7%; Score 8; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLLV 1141
|||||

Db 21 LLLALLV 28

RESULT 20
US-09-354-243B-28
Sequence 28, Application US/09354243B
Patent No. 6359117
GENERAL INFORMATION:
APPLICANT: Dumoutier, Laure
APPLICANT: Loubet, Jamila
APPLICANT: Renaud, Jean-Christophe
TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fa
TITLE OF INVENTION: (TIFFS)
FILE REFERENCE: LUD-5543.1
CURRENT APPLICATION NUMBER: US/09/354,243B
CURRENT FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: US09/178,973
PRIOR FILING DATE: 1998-10-26
NUMBER OF SEQ ID NOS: 29
SEQ ID NO 28
LENGTH: 179
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
US-09-354-243B-28

Query Match 0.7%; Score 8; DB 4; Length 179;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
Db 21 LLLALLV 28

RESULT 21
US-09-040-483-5
Sequence 5, Application US/09040483
Patent No. 6143867
GENERAL INFORMATION:
APPLICANT: Akerblom, Ingrid E.
TITLE OF INVENTION: NOVEL HUMAN EOSINOPHIL-
TITLE OF INVENTION: DERIVED BASIC PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,483
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/740,036
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0133 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 544241
US-09-040-483-5

Query Match 0.7%; Score 8; DB 4; Length 234;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 LLLALLV 1141
Db 4 LLLALLV 11

RESULT 22
US-08-485-618-103
Sequence 103, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6500 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-103

Query Match 0.7%; Score 8; DB 1; Length 494;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139

Db 442 GGLLLALL 449

RESULT 23

US-08-605-672-103
Sequence 103, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No.5817515el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27666/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-103

Query Match 0.7%; Score 8; DB 2; Length 494;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139
Db 442 GGLLLALL 449

RESULT 24
US-08-482-293A-103
Sequence 103, Application US/08482293A
Patent No. 5831029
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No.5831029el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,293A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27666/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-293A-103

Query Match 0.7%; Score 8; DB 2; Length 494;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLALL 1139
Db 442 GGLLLALL 449

RESULT 25
US-08-943-363-103
Sequence 103, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No.5837478el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 103:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-103

Query Match
Best Local Similarity 100.0%; DB 2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGILLLAL 1139
Db 442 GGILLLAL 449

RESULT 26
US-09-193-043-103
Sequence 103, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395el Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193,043
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286,889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 103
LENGTH: 494
TYPE: PRT
ORGANISM: rabbit
US-09-193-043-103

Query Match
Best Local Similarity 100.0%; DB 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGILLLAL 1139
Db 442 GGILLLAL 449

RESULT 27
US-09-688-307A-103

Sequence 103, Application US/09688307A
Patent No. 6432404
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6432404el Human Beta-2
FILE REFERENCE: 27866/36646
CURRENT APPLICATION NUMBER: US/09/688,307A
CURRENT FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 08/193,043
PRIOR FILING DATE: 1998-11-16
PRIOR APPLICATION NUMBER: 08/605,672
PRIOR FILING DATE: 1996-02-22
PRIOR APPLICATION NUMBER: 08/173,497
PRIOR FILING DATE: 1993-12-23
PRIOR APPLICATION NUMBER: 08/286,889
PRIOR FILING DATE: 1994-08-05
PRIOR APPLICATION NUMBER: 08/362,652
PRIOR FILING DATE: 1994-12-21
PRIOR APPLICATION NUMBER: 08/943,363
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 103
LENGTH: 494
TYPE: PRT
ORGANISM: rabbit
US-09-688-307A-103

Query Match
Best Local Similarity 100.0%; DB 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGILLLAL 1139
Db 442 GGILLLAL 449

RESULT 28
US-09-291-922-20
Sequence 20, Application US/09291922
Patent No. 6383776
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Hltz, Bill
APPLICANT: Kinney, Tony
APPLICANT: Tingey, Scott
TITLE OF INVENTION: Plant Sugar Transport Proteins
FILE REFERENCE: BB-1163
CURRENT APPLICATION NUMBER: US/09/291,922
CURRENT FILING DATE: 1999-04-14
EARLIER APPLICATION NUMBER: 60/083,044
EARLIER FILING DATE: April 24, 1998
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Microsoft Office 97
SEQ ID NO 20
LENGTH: 513
TYPE: PRT
ORGANISM: Zea mays
US-09-291-922-20

Query Match
Best Local Similarity 100.0%; DB 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1136 LLALLVFC 1143
Db 192 LLALLVFC 199

RESULT 29
US-09-409-648-7
Sequence 7, Application US/09409648

Patent No. 6368797
; GENERAL INFORMATION:
; APPLICANT: Keith Schappert
; TITLE OF INVENTION: METHODS FOR TREATING OR IDENTIFYING A
; TITLE OF INVENTION: SUBJECT AT RISK FOR A NEUROLOGICAL DISEASE BY DETERMINING
; TITLE OF INVENTION: THE PRESENCE OF A VARIANT GPIIb AND/OR VARIANT GPIIb ALLELE
; FILE REFERENCE: 08523/015002
; CURRENT APPLICATION NUMBER: US/09/409,648
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 60/102,624
; EARLIER FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 7
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-409-648-7

Query Match 0.7%; Score 8; DB 4; Length 1039;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 VLGGLLLL 1137
Db 1004 VLGGLLLL 1011

RESULT 30
US-09-409-648-8
; Sequence 8, Application US/09409648
; Patent No. 6368797
; GENERAL INFORMATION:
; APPLICANT: Keith Schappert
; TITLE OF INVENTION: METHODS FOR TREATING OR IDENTIFYING A
; TITLE OF INVENTION: SUBJECT AT RISK FOR A NEUROLOGICAL DISEASE BY DETERMINING
; TITLE OF INVENTION: THE PRESENCE OF A VARIANT GPIIb AND/OR VARIANT GPIIb ALLELE
; FILE REFERENCE: 08523/015002
; CURRENT APPLICATION NUMBER: US/09/409,648
; CURRENT FILING DATE: 1999-10-01
; EARLIER APPLICATION NUMBER: 60/102,624
; EARLIER FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 8
; LENGTH: 1039
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-409-648-8

Query Match 0.7%; Score 8; DB 4; Length 1039;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 VLGGLLLL 1137
Db 1004 VLGGLLLL 1011

RESULT 31
5196511-2
; Patent No. 5196511
; APPLICANT: PLOM, EDWARD F., D'SOUZA, STANLEY E.
; GINSBERG, MARK H.
; TITLE OF INVENTION: PEPTIDES AND ANTIBODIES THAT INHIBIT
; INTERLIN-LIGAND BINDING
; NUMBER OF SEQUENCES: 31
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/444,777
; FILING DATE: 01-DEC-1989
; SEQ ID NO: 2
; LENGTH: 1039
5196511-2

Query Match 0.7%; Score 8; DB 6; Length 1039;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1130 VLGGLLLL 1137
Db 1004 VLGGLLLL 1011

RESULT 32
US-08-286-889-37
; Sequence 37, Application US/08286889
; Patent No. 5470953
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Mich
; TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit.
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESS: 233 South Wacker Drive, 6300 Sear Tower
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,889
; FILING DATE:

CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: P38,659
; REFERENCE/DOCKET NUMBER: 27866/32168
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-889-37

Query Match 0.7%; Score 8; DB 1; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLALL 1139
Db 1100 GGLLLALL 1107

RESULT 33
US-08-485-618-37
; Sequence 37, Application US/08485618
; Patent No. 5728533
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-37
Query Match 0.7%; Score 8; DB 1; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1132 GGLLLAL 1139
DB 1100 GGLLLAL 1107
RESULT 34
US-08-362-652-37
Sequence 37, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,652
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32391
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-652-37
Query Match 0.7%; Score 8; DB 1; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1132 GGLLLAL 1139
DB 1100 GGLLLAL 1107
RESULT 35
US-08-605-672-37
Sequence 37, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5817515e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448

TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-37

Query Match 0.7%; Score 8; DB 2; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1100 GGLLLAL 1107

RESULT 36
US-08-482-293A-37
Sequence 37, Application US/08482293A
Patent No. 5831029
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,293A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-293A-37

Query Match 0.7%; Score 8; DB 2; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139

Db 1100 GGLLLAL 1107

RESULT 37
US-08-943-363-37
Sequence 37, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1151 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-37

Query Match 0.7%; Score 8; DB 2; Length 1151;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1100 GGLLLAL 1107

RESULT 38
US-09-193-043-37
Sequence 37, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395e1 Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193,043

```

; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-37

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 1151;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1100 GGLLLAL 1107

RESULT 39
US-09-688-307A-37
; Sequence 37, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 1151
; TYPE: PRT
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 361
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 464
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 466
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 506
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 1117
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 1118
```

```

; OTHER INFORMATION: Xaa = any or unknown amino acid
US-09-688-307A-37

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 1151;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1100 GGLLLAL 1107

RESULT 40
US-08-476-062A-43
; Sequence 43, Application US/08476062A
; Patent No. 5877275
; GENERAL INFORMATION:
; APPLICANT: Arimaout, M. Amin
; TITLE OF INVENTION: CONTROLLING CELLULAR IMMUNE/INFLAMMATORY
; TITLE OF INVENTION: RESPONSES WITH BETA2 INTEGRINS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,062A
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,081
; FILING DATE: 21-MAR-1994
; APPLICATION NUMBER: 07/637,830
; FILING DATE: 04-JAN-1991
; APPLICATION NUMBER: 07/539,842
; FILING DATE: 18-JUN-1990
; APPLICATION NUMBER: 07/212,573
; FILING DATE: 28-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00786/068003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-476-062A-43

Query Match
Best Local Similarity 100.0%; Score 8; DB 2; Length 1152;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1115 GGLLLAL 1122

RESULT 41
PCT-US96-01314-43
```

Sequence 43, Application PC/TUS9601314
GENERAL INFORMATION:
APPLICANT: M. Amin Arnaout
TITLE OF INVENTION: METHODS FOR IDENTIFYING INTEGRIN
TITLE OF INVENTION: ANTAGONISTS
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/01314
FILING DATE: 30-JAN-96
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/380,167
FILING DATE: 30-JAN-95
ATTORNEY/AGENT INFORMATION:
NAME: John W. Freeman
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00786/267001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 1152
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-01314-43

Query Match 0.7%; Score 8; DB 5; Length 1152;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
Db 1115 GGLLLAL 1122

RESULT 42
5424399-2
Patent No. 5424399
APPLICANT: ARNAOUT, M. AMIN
TITLE OF INVENTION: HUMAN CR3a/b HETERODIMERS
NUMBER OF SEQUENCES: 12
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/78,871
FILING DATE: 16-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 539,842
FILING DATE: 18-JUN-1990
APPLICATION NUMBER: 212,573
FILING DATE: 28-JUN-1988
SEQ ID NO: 2;
LENGTH: 1152
5424399-2

Query Match 0.7%; Score 8; DB 6; Length 1152;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||

Db 1115 GGLLLAL 1122

RESULT 43
US-08-173-497-3
Sequence 3, Application US/08173497
Patent No. 5437958
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No. 5437958e1 Human 2 Integrin Alpha
TITLE OF INVENTION: Subunit
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 S. Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/173,497
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5437958and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31363
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-497-3

Query Match 0.7%; Score 8; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
Db 1116 GGLLLAL 1123

RESULT 44
US-08-286-889-3
Sequence 3, Application US/08286889
Patent No. 5470953
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Mich
TITLE OF INVENTION: No. 5470953e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Seear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: P38,659
REFERENCE/DOCKET NUMBER: 27866/32168
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-889-3

Query Match 0.7%; Score 8; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1116 GGLLLAL 1123

RESULT 45
US-08-485-618-3
Sequence 3, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659

REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-3

Query Match 0.7%; Score 8; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1116 GGLLLAL 1123

RESULT 46
US-08-362-652-3
Sequence 3, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,652
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32391
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-652-3

Query Match 0.7%; Score 8; DB 1; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1132 GGLLLAL 1139
1116 GGLLLAL 1123

RESULT 47
US-08-605-672-3
; Sequence 3, Application US/08605672
; Patent No. 5817515
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/605,672
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1153 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-605-672-3

Query Match 0.7%; Score 8; DB 2; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1132 GGLLLAL 1139
1116 GGLLLAL 1123

RESULT 48
US-08-482-293A-3
; Sequence 3, Application US/08482293A
; Patent No. 5831029
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael

APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,293A
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/173,497
; FILING DATE: 23-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,889
; FILING DATE: 5-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,652
; FILING DATE: 21-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/32684
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1153 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-482-293A-3

Query Match 0.7%; Score 8; DB 2; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1132 GGLLLAL 1139
1116 GGLLLAL 1123

RESULT 49
US-08-943-363-3
; Sequence 3, Application US/08943363
; Patent No. 5837478
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sear Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1153 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-3

Query Match 0.7%; Score 8; DB 2; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1116 GGLLLAL 1123

RESULT 50
US-09-193-043-3
Sequence 3, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395el Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193,043
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286,889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1153
TYPE: PRT
ORGANISM: Homo sapiens
US-09-193-043-3

Query Match 0.7%; Score 8; DB 4; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139

Db 1116 GGLLLAL 1123

RESULT 51
US-09-688-307A-3
Sequence 3, Application US/09688307A
Patent No. 6432404
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6432404el Human Beta-2
FILE REFERENCE: 27866/36646
CURRENT APPLICATION NUMBER: US/09/688,307A
CURRENT FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 09/193,043
PRIOR FILING DATE: 1998-11-16
PRIOR APPLICATION NUMBER: 08/605,672
PRIOR FILING DATE: 1996-02-22
PRIOR APPLICATION NUMBER: 08/173,497
PRIOR FILING DATE: 1993-12-23
PRIOR APPLICATION NUMBER: 08/286,889
PRIOR FILING DATE: 1994-08-05
PRIOR APPLICATION NUMBER: 08/362,652
PRIOR FILING DATE: 1994-12-21
PRIOR APPLICATION NUMBER: 08/943,363
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1153
TYPE: PRT
ORGANISM: Homo sapiens
US-09-688-307A-3

Query Match 0.7%; Score 8; DB 4; Length 1153;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1116 GGLLLAL 1123

RESULT 52
US-08-286-889-46
Sequence 46, Application US/08286889
Patent No. 5470953
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Mich
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: P38,659
REFERENCE/DOCKET NUMBER: 27866/32168
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-889-46

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 1104 GGLLLAL 1111

RESULT 53
US-08-485-618-46
Sequence 46, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-46

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 1104 GGLLLAL 1111

RESULT 54
US-08-362-652-46
Sequence 46, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,652
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32391
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-652-46

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1155;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 1104 GGLLLAL 1111

RESULT 55
US-08-605-672-46
Sequence 46, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica

TITLE OF INVENTION: No. 5817515e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-46

Query Match 0.7%; Score 8; DB 2; Length 1155;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 1104 GGLLLAL 1111

RESULT 56
US-08-482-293A-46
Sequence 46, Application US/08482293A
Patent No. 5831029
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5831029e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,293A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-482-293A-46

Query Match 0.7%; Score 8; DB 2; Length 1155;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
DB 1104 GGLLLAL 1111

RESULT 57
US-08-943-363-46
Sequence 46, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478e1 Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1155 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-46

Query Match 0.7%; Score 8; DB 4; Length 1155;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
Db 1104 GGLLLAL 1111

RESULT 58
US-09-193-043-46
Sequence 46, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395el Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193,043
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286,889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 46
LENGTH: 1155
TYPE: PRT
ORGANISM: Mus musculus
US-09-193-043-46

Query Match 0.7%; Score 8; DB 4; Length 1155;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
Db 1104 GGLLLAL 1111

RESULT 59
US-09-688-307A-46
Sequence 46, Application US/09688307A
Patent No. 6432404
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6432404el Human Beta-2
FILE REFERENCE: 27866/36646
CURRENT APPLICATION NUMBER: US/09/688,307A
CURRENT FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 09/193,043
PRIOR FILING DATE: 1998-11-16

PRIOR APPLICATION NUMBER: 08/605,672
PRIOR FILING DATE: 1996-02-22
PRIOR APPLICATION NUMBER: 08/173,497
PRIOR FILING DATE: 1993-12-23
PRIOR APPLICATION NUMBER: 08/286,889
PRIOR FILING DATE: 1994-08-05
PRIOR APPLICATION NUMBER: 08/362,652
PRIOR FILING DATE: 1994-12-21
PRIOR APPLICATION NUMBER: 08/943,363
PRIOR FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 46
LENGTH: 1155
TYPE: PRT
ORGANISM: Mus musculus
US-09-688-307A-46

Query Match 0.7%; Score 8; DB 4; Length 1155;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
Db 1104 GGLLLAL 1111

RESULT 60
US-08-485-618-53
Sequence 53, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESSES:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Seear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-53

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 61

US-08-485-618-55
Sequence 55, Application US/08485618
Patent No. 5728533
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5728533el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,618
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32797
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-618-55

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 62.

US-08-362-652-53
Sequence 53, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

US-08-362-652-53
Sequence 53, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-08-362-652-55
Sequence 55, Application US/08362652
Patent No. 5766850
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5766850el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole; Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 1161;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,652
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32391
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-652-55

Query Match 0.7%; Score 8; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 64
US-08-605-672-53
Sequence 53, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:

NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-53

Query Match 0.7%; Score 8; DB 2; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1132 GGLLLAL 1139
Db 1110 GGLLLAL 1117

RESULT 65
US-08-605-672-55
Sequence 55, Application US/08605672
Patent No. 5817515
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5817515el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 103
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/605,672
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-605-672-55

Query Match 0.7%; Score 8; DB 2; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
 |||||
 Db 1110 GGLLLAL 1117

RESULT 66

US-08-482-293A-53
 ; Sequence 53, Application US/08482293A
 ; Patent No. 5831029
 ; GENERAL INFORMATION:
 ; APPLICANT: Gallatin, W. Michael
 ; APPLICANT: Van der Vieren, Monica
 ; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
 ; NUMBER OF SEQUENCES: 103
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/482,293A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/173,497
 ; FILING DATE: 23-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/286,889
 ; FILING DATE: 5-AUG-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,652
 ; FILING DATE: 21-DEC-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Williams Jr., Joseph A.
 ; REGISTRATION NUMBER: 38,659
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-474-6300
 ; TELEFAX: 312-474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 53:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1161 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-482-293A-53

Query Match 0.7%; Score 8; DB 2; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
 |||||
 Db 1110 GGLLLAL 1117

RESULT 67
 US-08-482-293A-55
 ; Sequence 55, Application US/08482293A
 ; Patent No. 5831029

GENERAL INFORMATION:
 ; APPLICANT: Gallatin, W. Michael
 ; APPLICANT: Van der Vieren, Monica
 ; TITLE OF INVENTION: No. 5831029el Human 2 Integrin Alpha Subunit
 ; NUMBER OF SEQUENCES: 103
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/482,293A
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/173,497
 ; FILING DATE: 23-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/286,889
 ; FILING DATE: 5-AUG-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,652
 ; FILING DATE: 21-DEC-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Williams Jr., Joseph A.
 ; REGISTRATION NUMBER: 38,659
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-474-6300
 ; TELEFAX: 312-474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 55:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1161 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-482-293A-55

Query Match 0.7%; Score 8; DB 2; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
 |||||
 Db 1110 GGLLLAL 1117

RESULT 68
 US-08-943-363-53
 ; Sequence 53, Application US/08943363
 ; Patent No. 5837478
 ; GENERAL INFORMATION:
 ; APPLICANT: Gallatin, W. Michael
 ; APPLICANT: Van der Vieren, Monica
 ; TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
 ; NUMBER OF SEQUENCES: 114
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive, 6300 Sear Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-53

Query Match 0.7%; Score 8; DB 2; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
DB 1110 GGLLLAL 1117

RESULT 69
US-08-943-363-55
Sequence 55, Application US/08943363
Patent No. 5837478
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Michael
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 5837478el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,363
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,889
FILING DATE: 5-AUG-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/362,652
FILING DATE: 21-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/32684
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-363-55

Query Match 0.7%; Score 8; DB 2; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
DB 1110 GGLLLAL 1117

RESULT 70
US-09-193-043-53
Sequence 53, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395el Human 2
FILE REFERENCE: 27866/35004
CURRENT FILING DATE: 1998-11-16
EARLIER APPLICATION NUMBER: 08/173,497
EARLIER FILING DATE: 1993-12-23
EARLIER APPLICATION NUMBER: 08/286,889
EARLIER FILING DATE: 1994-08-05
EARLIER APPLICATION NUMBER: 08/362,652
EARLIER FILING DATE: 1994-12-21
EARLIER APPLICATION NUMBER: 08/943,363
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 53
LENGTH: 1161
TYPE: PRT
ORGANISM: Mus musculus
US-09-193-043-53

Query Match 0.7%; Score 8; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1132 GGLLLAL 1139
|||||
DB 1110 GGLLLAL 1117

RESULT 71
US-09-193-043-55
Sequence 55, Application US/09193043
Patent No. 6251395
GENERAL INFORMATION:
APPLICANT: Gallatin, Michael W.
APPLICANT: Van der Vieren, Monica
TITLE OF INVENTION: No. 6251395el Human 2
FILE REFERENCE: 27866/35004
CURRENT APPLICATION NUMBER: US/09/193,043

```

; CURRENT FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: 08/173,497
; EARLIER FILING DATE: 1993-12-23
; EARLIER APPLICATION NUMBER: 08/286,889
; EARLIER FILING DATE: 1994-08-05
; EARLIER APPLICATION NUMBER: 08/362,652
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: 08/943,363
; EARLIER FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-193-043-55

Query Match      0.7%; Score 8; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1132 GGLLLAL 1139
DB      1110 GGLLLAL 1117

RESULT 72
US-09-688-307A-53
; Sequence 53, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-688-307A-53

Query Match      0.7%; Score 8; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1132 GGLLLAL 1139
DB      1110 GGLLLAL 1117

RESULT 73
US-09-688-307A-55
; Sequence 55, Application US/09688307A
; Patent No. 6432404
; GENERAL INFORMATION:
; APPLICANT: Gallatin, Michael W.
; APPLICANT: Van der Vieren, Monica

```

```

; TITLE OF INVENTION: No. 6432404el Human Beta-2
; FILE REFERENCE: 27866/36646
; CURRENT APPLICATION NUMBER: US/09/688,307A
; CURRENT FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 09/193,043
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 08/605,672
; PRIOR FILING DATE: 1996-02-22
; PRIOR APPLICATION NUMBER: 08/173,497
; PRIOR FILING DATE: 1993-12-23
; PRIOR APPLICATION NUMBER: 08/286,889
; PRIOR FILING DATE: 1994-08-05
; PRIOR APPLICATION NUMBER: 08/362,652
; PRIOR FILING DATE: 1994-12-21
; PRIOR APPLICATION NUMBER: 08/943,363
; PRIOR FILING DATE: 1997-10-03
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 1161
; TYPE: PRT
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 474
; OTHER INFORMATION: Xaa = any or unknown amino acid
; NAME/KEY: misc_feature
; LOCATION: 1128
; OTHER INFORMATION: Xaa = any or unknown amino acid
US-09-688-307A-55

Query Match      0.7%; Score 8; DB 4; Length 1161;
Best Local Similarity 100.0%; Pred. No. 72;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      1132 GGLLLAL 1139
DB      1110 GGLLLAL 1117

RESULT 74
US-08-173-497-4
; Sequence 4, Application US/08173497
; Patent No. 5437958
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Van der Vieren, Monica
; TITLE OF INVENTION: No. 5437958el Human 2 Integrin Alpha
; TITLE OF INVENTION: Subunit
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/173,497
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5437958and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31363
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448

```

Search completed: July 16, 2003, 08:34:39
Job time : 23 secs

TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1163 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-173-497-4

Query Match 0.7%; Score 8; DB 1; Length 1163;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1132 GGLLLAL 1139
|||
Db 1115 GGLLLAL 1122

RESULT 75
US-08-286-889-4
Sequence 4, Application US/08286889
Patent No. 5470953
GENERAL INFORMATION:
APPLICANT: Gallatin, W. Mich
APPLICANT: Van der Vlieten, Monica
TITLE OF INVENTION: No. 5470953el Human 2 Integrin Alpha Subunit
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sear Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,889
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/173,497
FILING DATE: 23-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: P38,659
REFERENCE/DOCKET NUMBER: 27866/32168
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1163 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-889-4

Query Match 0.7%; Score 8; DB 1; Length 1163;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1132 GGLLLAL 1139
|||
Db 1115 GGLLLAL 1122

THIS PAGE BLANK (USPTO)